Investigation on the Structure of CpG Methylated DNA

by

Carmel M. Reilly

Presented for:

The Degree of Doctor of Philosophy (PhD)
The University of Edinburgh
2002

Table of Contents

Table of Contents ..... (I) →(III)
Table of Figures IV
Table of Tables VI
Abbreviations VII
Declaration of Intent IX
Acknowledgements X
Abstract XI

Chapter 1: Introduction
1.1 The Significance of CpG Methylation in the cell: An overview 1
1.2 The Effects of CpG methylation on DNA Bending 8
1.3 DNA Conformation: A-, B- and Z-DNA 15
1.4 The Effects of CpG Methylation on DNA Conformation 21
1.4.1 Z-DNA 21
1.4.2 A-DNA 28
1.5 Methyl-CpG binding Proteins and DNA recognition: A role for DNA structure? 31
1.6 Methyl-CpG binding proteins: the family 33
1.7 Aims of Thesis 46

Chapter 2: Materials and Methods
2.1 Materials 47
2.1.1 List of Suppliers 47
2.1.2 Stock Solutions List 50
2.2 Methods 59
2.2.1 Preparation of Competent Cells 60
2.2.2 Transformation and growth of plasmid DNA in bacterial host strains 60
2.2.3 Agarose Gel Electrophoresis 61
2.2.4 Polyacrylamide Gel Electrophoresis 61
2.2.5 Staining of Agarose and PAGE gels 61
2.2.6 Medium scale growth (midiprep) of plasmid DNA 62
2.2.7 Small-scale growth (miniprep) of plasmid DNA 63
2.2.8 Amplification of DNA by the Polymerase Chain Reaction (PCR) 63
2.2.9 Dephosphorylation pf 5'-ends of dsDNA for radioactive labelling 64
2.2.10 Preparation of radiolabelled marker DNA 64
2.2.11 Radioactive labelling of substrate DNA for Benzonase, DnaseI and Footprinting analyses 66
2.2.12 Preparation of Running of denaturing PAGE/Urea Sequencing gels 66
2.2.13 Drying and Exposure of PAGE/UREA gels 67
2.2.14 SDS-PAGE for the denaturation and resolution of proteins 67
2.2.15 Induction procedure for GST-tagged MBD protein 68
2.2.16 Fast Method to test IPTG induced cell cultures for effective protein expression 69

2.2.17 GST purification using Glutathione Sepharose 4B beads *70*
2.2.18 Expression of pET6hMBD; a histidine tagged MBD protein *70*
2.2.19 Dialysis of the dissolved pellets from pET6hMBD expression *71*
2.2.20 Purification of the MBD protein using a Fractogel EMD-SO_3^-650 column ... *72*
2.2.21 Purification of the Fractogel purified histidine tagged MBD protein over a Ni^{2+} charged affinity column *72*
2.2.22 Purification of supernatant fraction from pET6hMBD expression *73*
2.2.23 Annealing of complementary ssDNA oligomers *74*
2.2.24 Bandshift analysis using bacterial cell purified proteins *74*
2.2.25 Protocol for Benzonase Digestions of Radiolabelled DNA *75*
2.2.26 Protocol for DNaseI Digestions of Radiolabelled DNA *75*
2.2.27 Protocol for Footprinting analysis of Radiolabelled DNA *79*
2.2.28 Western Blotting Procedure for histidine tagged or GST tagged proteins .. *79*

Chapter 3:Results (I)

3.1 Devising a method to detect the presence of A-DNA in CpG methylated DNA *80*
3.1.1 Introduction *80*
3.1.2 Results *83*
3.1.2.1 Can the Benzonase assay detect A-form DNA in CpG methylated DNA? *83*
3.1.2.2 Benzonase analysis on the Amp^R gene of pBR322: the effect of methylation *84*
3.1.2.3 DNaseI analysis on the Amp^R gene of pBR322: the effect of methylation *91*
3.1.2.4Benzonase analysis of a fragment of the pGEM9Zf(-) plasmid vector ... *97*
3.1.2.5 Refining the technique: The use of Urea to promote Benzonase specificity *98*
3.1.2.6 DNA methylation and globin gene regulation *106*
3.1.2.6.1 Preparation of human α_2-lobin gene fragments for benzonase analysis *107*
3.1.2.6.2 Benzonase analysis of the human α_2-globin gene (SssI methylated) *108*
3.1.3 Summary *114*

Chapter 4: Results (II)

4.1 Analysis of the chicken β^A-globin gene *118*
4.1.1 Introduction *118*
4.1.2 Results *122*
4.1.2.1 β^A-globin gene fragments *122*
4.1.2.2 Benzonase analysis of β^A-globin wild-type (LE) fragments (+ strand) *125*
4.1.2.3 Benzonase analysis of β^A-globin wild-type (LE) fragments (- strand) . *129*
4.1.2.4 Benzonase analysis of β^A-globin LE and mutant 1 (m1) fragments *137*
4.1.2.5 Benzonase analysis of β^A-globin LE and mutant 2 (m2) fragments *139*
4.1.2.6 Benzonase analysis of β^A-globin mutant 3 (m3) fragments *142*
4.1.2.7 Benzonase analysis of β^A-globin mutant 23 (m23) fragments *145*
4.1.2.8 Benzonase analysis of β^A-globin mutant 25 (m25) fragments *149*
4.1.2.9 Benzonase analysis of β^A-globin LE and mutant 8a (m8a) fragments *151*
4.2 Summary *151*

Chapter 5: Results (III)
5.1 Bandshift experiments with purified methyl binding proteins *153*
5.1.1 Design of oligos for bandshift analyses *153*
5.1.2 Purification of the MBD domain of MeCP2 ..... *153*
5.1.3 Bandshift experiments with the purified MBD domain of MeCP2 ..... *157*
5.1.3.1 The MeCP2 MBD domain and the A-dnaM oligo *157*
5.1.3.2 The MeCP2 MBD domain and the A-dnaU oligo *158*
5.1.3.3 The MeCP2 MBD domain and the A-dnaU oligo and A-dnaM oligos with competitor DNA *158*
5.1.3.4 The MeCP2 MBD domain and the B-dnaU oligo *158*
5.1.3.5 The MeCP2 MBD domain and the BgloU oligo and BgloM oligos *163*
5.1.3.6 Bandshift experiments with the purified xMBD1 protein *163*
5.1.4 Footprinting analysis of the β^A-globin gene promoter region with purified MBD protein *168*
5.1.5 Summary *170*

Chapter 6: Discussion
6.1 The Benzonase Enzyme: History and specificity for A-form DNA *173*
6.2 CpG Methylation Potentiates the formation of A-form DNA *176*
6.3 How can methylation induced changes in DNA structure influence where and when methyl-binding proteins bind? *196*

Chapter 7: Appendix
7.1 Benzonase analysis of a DNA fragment from pBR322 ..... *202*
7.2 Benzonase analysis of a DNA fragment from the human H19 gene *202*
7.3 Benzonase analysis of a DNA fragment from a xMeCP2 ChIP assay *203*
7.4 The Integrase Assay: Expression of HIV-1 integrase *203*
7.5 Expression and Purification of the HIV-1 integrase protein *208*
7.6 Cleavage HIV-1 integrase histidine tag with Thrombin *211*
7.7 Removal of Thrombin with a Benzamidine Sepharose 6B column *211*
7.8 Concentration of eluted protein on a Fractogel EMD-SO_3^- column *212*
7.9 Western blotting to detect presence of HIV-1 integrase *212*
7.10 The Integrase Assay: Attempts at DNA structural analysis *212*
7.11 Creation of a CpG methylation specific restriction enzyme *215*

References ***217-236***

Table of Figures

Figure 1.1 The conversion of cytosine to 5-methyl-cytosine..... 2
Figure 1.2 Higher order levels of chromatin compaction..... 4
Figure 1.3 A model for RNA mediated silencing in plants..... 6
Figure 1.4 B-DNA..... 17
Figure 1.5 A-DNA..... 18
Figure 1.6 Z-DNA..... 19
Figure 1.7 A-form DNA successfully competes with methylated DNA for MeCP1 binding..... 32
Figure 1.8 *A* The methyl binding proteins..... 34
Figure 1.8 B Sequence alignment of the MBD from the MBD proteins..... 34
Figure 1.9A The structure of the MBD domain from MeCP2..... 36
Figure 1.9 *B* The solution structure of the MBD domain of MBD1..... 36
Figure 1.10 Mutations in MeCP2 found in Rett Syndrome patients..... 38
Figure 1.11 The methylation machinery: interacting components..... 45

Figure 3.1 The Integrase assay: An outline..... 81
Figure 3.2 Benzonase digest of pBR322 plasmid..... 85
Figure 3.3 The Benzonase assay: An outline..... 87
Figure 3.4 The sequence of pBR322 used for Benzonase analysis..... 88
Figure 3.5 *A* Benzonase digestion of pBR322 fragments..... 89
Figure 3.5 *B* Magnified view of Image 3.5 A..... 90
Figure 3.5 C Peak determination scan of figure 3.5 A..... 90
Figure 3.6 *A* DNaseI digestion of pBR322 fragments..... 93
Figure 3.6 *B* Magnified view of Image 3.6 A. DNaseI analysis on pBR322..... 94
Figure 3.6 C Peak determination scan of selected regions of the DNaseI digestion..... 95 & 96
Figure 3.7 The fragment from pGEM9Zf (-) used in Benzonase analysis..... 99
Figure 3.8 A Benzonase digest on pGEM9Zf(-) fragments..... 100
Figure 3.8 B Peak determination scan of the Benzonase analysis performed in figure 3.8 A..... 101
Figure 3.9 Benzonase digest of pBR322 fragments in altered conditions..... 103
Figure 3.10 A Benzonase digest of pGEM9Zf(-) fragments in altered conditions..... 104
Figure 3.10 B Peak determination scan of the Benzonase digestion assay shown in figure 3.10 A..... 105
Figure 3.11 The human alpha-2 globin gene sequence..... 109
Figure 3.12 *A* Methylation protection assay of pUC9 plasmid containing the human alpha-2 globin gene..... 110
Figure 3.13 A Benzonase digest of the human alpha-2 globin gene..... 111
Figure 3.13 B, C, D Peak determination scans of selected regions of the Benzonase assay shown in figure 3.14 A..... 112
Figure 3.14 Benzonase analysis of the human alpha-2 globin gene..... 113

Figure 4.1 CpG methylation alters nucleosome positioning..... 120
Figure 4.2 The chicken β-globin locus..... 123

Figure 4.3 A Benzonase analysis of the β^A-globin wild-type LE fragments (+ strand) 127
Figure 4.3 B Magnified view of the highlighted section from figure 4.3 A 128
Figure 4.3 C Peak determination scan using Aida™ software 128
Figure 4.4 A Benzonase analysis of β^A-globin wild-type fragments (- strand) 131
Figure 4.4 B Enlarged region from figure 4.6 A 132
Figure 4.4 C Peak determination scan of the Benzonase ddigestion analysis shown in figure 4.4 A 132
Figure 4.5 A DNaseI analysis of β^A-globin promoter fragments 133
Figure 4.5 B Magnified view of the DNaseI analysis shown in figure 4.5 A 136
Figure 4.6 A Monomer extension: A method to identify nucleosome positioning sites in DNA 134
Figure 4.6 B Monomer extension analysis of β^A-globin promoter fragments .. 135
Figure 4.6 C Magnified view of the nucleosome positioning analysis shown in figure 4.6 B 136
Figure 4.7 Benzonase analysis of wild-type LE fragments compared to m1 138
Figure 4.8 A Benzonase digestion of wild-type LE fragments compared to m2 141
Figure 4.8 B Peak determination scan of the CpG triplet region in the mutant m2 141
Figure 4.9 A Benzonase digestion of wild-type LE fragments compared to m3 143
Figure 4.9 B Peak determination scans of the CpG triplet region of m3 143
Figure 4.9 C Peak determination scan of the CpG triplet region of m3 144
Figure 4.10 I Benzonase digestion of mutants m23 and m25 fragments 146
Figure 4.10 B & C Peak determination scans of the CpG triplet region of mutants m23 and m25 147
Figure 4.11 ABenzonase digest of wild-type LE fragments compared to mutant 8a 150
Figure 4.11 B Peak determination scans of the m8a mutant 150

Figure 5.1 Purification of the MBD domain of MeCP2 using a cation exchange column 155
Figure 5.2 Purification of the MBD domain from MeCP2 using a Nickel affinity column 156
Figure 5.3 Bandshift analysis with purified MBD protein and A-dnaM oligo 159
Figure 5.4 Bandshift analysis of purified MBD domain of MeCP2 with the A-dnAU oligomer 160
Figure 5.5 Bandshift analysis of purified MBD protein and the A-dnaU and A-dnaM oligos in the presence of competitor DNA 161
Figure 5.6 Bandshift analysis of purified MBD protein with the B-dnaU oligo162
Figure 5.7 Bandshift analysis of purified MBD protein with the A-dnaU and B-dnaU oligo 164
Figure 5.8 Bandshift analysis of purified MBD protein with B-gloU oligos 165
Figure 5.9 Bandshift analysis of purified MBD protein with B-gloU and B-gloM oligos 166

Figure 5.10 Bandshift analysis of the xMBD1 protein with the A-dnaU and A-dnaM oligomers 167
Figure 5.11 DNaseI footprint analysis of methylated β^A-globin LE fragments with increasing amounts of MBD and constant amounts of DNaseI 169

Figure 6.1 The chromosomal organisation of the human globin genes.....181
Figure 6.2 How can methylation induced changes in DNA structure influence where and when methyl-binding proteins bind? 196

Figure 7.1 Benzonase analysis of the reverse strand of a pBR322 fragment....204
Figure 7.2 Benzonase analysis of a region of the human H19 gene.....205
Figure 7.3 Benzonase analysis of a ChIP-MeCP2 pull-down fragment.....206
Figure 7.4 Expression of HIV-1 integrase protein.....209
Figure 7.5 Purification of a histidine tagged HIV-integrase protein.....210
Figure 7.6 Removal of Thrombin protease with a Benzamidine 6B column....213
Figure 7.7 Concentration of HIV-1 integrase..... 214
Figure 7.8 HIV-1 integrase expression.....214
Figure 7.9 Construction of a methylCpG specific nuclease.....214

Table of Tables

Table 1.1 Structural features of idealised A-, B- and Z-DNA tertiary structures16
Table 1.2 Crystal structures of alternating oligonucleotide sequences..... 22
Table 2.1 List of commonly used bacterial cells and plasmids 59
Table 2.2 Primers for PCR generation of DNA fragments for structural analysis 65
Table 2.3 Deenaturing PAGE/Urea gel mix recipies..... 66
Table 2.4 Separating gels for SDS-PAGE analysis 68
Table 2.5 The Benzonase Assay 76
Table 2.6 DNaseI analysis of DNA fragments 77
Table 2.7 DNaseI footprinting Protocol 78
Table 4.1 Site-directed mutagenesis in the β-^{A}-globin CpG triplet 121
Table 5.1 DNA oligos used for Bandshift analysis 154
Table 6.1 Benzonase and DNaseI digestion: Preferred cleavage sites at CpG methylated DNA sites..... 179
Table 6.2 Tabulated information for Benzonase, DNaseI and nucleosome displacement analyses.....185

Abbreviations

A ..... Adenine; Amps; Absorbance Units
aa ..... amino acid
$A_{260/280}$ nm ..... Absorbance at 260 nm
Ab ..... Antibody
Ac ..... Acetylated
A-DNA ..... A-form (DNA in a right-handed superhelix)
APS ..... Ammonium Persulphate
ATP ..... Adenosine Triphosphate
BAP ..... Bacterial alkaline phosphatase
B-DNA ..... B-form (DNA in a typical Watson-Crick superhelix)
B-gal ..... B-galactosidase
bp ..... Base pair
BSA ..... Bovine Serum Albumin
BZase ..... Benzon Nuclease
C ..... Cytosine
cfu ..... Colony forming unit
Ci ..... Curie
CIAP ..... Calf Intestinal Alkaline Phosphotase
CpG ..... Cytosine-guanine dinucleotide pair
cpm ..... Counts per minute
Da ..... Dalton
dATP ..... Deoxyadenosine Triphosphate
dCTP ..... Deoxycytidine triphosphate
ddH_2O ..... double-distilled H_20
DEAE ..... Diethylaminoethyl
DMSO ..... Dimethyl sulphoxide
DNA ..... Deoxyribonucleic acid
DNAseI ..... Deoyribonuclease I
dNTP ..... Deoxyribonucleoside triphosphate
DTT ..... Dithiothreitol
EDTA ..... Ethylenediaminetetracetic acid
E-DNA ..... E-form (stable superhelical intermediate bet. A-and B-form DNA)
ELISA ..... Enzyme linked immunosorbent Assay
EST ..... Expressed Sequence Tag
EtOH ..... Ethanol
EtBr ..... Ethidium Bromide
FCS ..... Foetal Calf Serum
g ..... Gravity
γ^{32}P-ATP ..... gamma (tertiary) phosphate labelled-ATP
HAT ..... Histone Acetyltransferase
HDAC ..... Histone Deacetylase
HeLa ..... Henrietta Laxx cells (immortal cell line)
IPTG Isopropyl-1-thio-B-D-galactoside
kb ..... Kilobase Pair
kDa ..... KiloDalton
MBD ..... Methyl binding domain

MeCP ..... Methyl CpG binding protein
me-CpG ..... Methylated-CpG dinucleotide pair
mins ..... Minutes
Mnase ..... Micrococcal Nuclease
M_r ..... Relative Molecular Weight
N ..... Nucleotide
OD_{260} ..... Optical Density at 260 nm
Oligo ..... Oligonucleotide
ORF ..... Open Reading Frame
PAGE ..... Polyacrylamide gel electrophoresis
PBS ..... Phosphate Buffered Saline
PCR ..... Polyacrylamide Chain Reaction
PEG ..... Polyethylene glycol
Pfu ..... Plaque forming unit
pI ..... Isoelectric point
PMSF ..... Phenylmethylsulphonyl fluoride
PNK ..... Polynucleotide Kinase
PVDF ..... Polyvinylidene difluoride
RNA ..... Ribonucleic acid
Rnase ..... Ribonuclease
rRNA ..... Ribosomal ribonucleic acid
rpm ..... revolutions per minute
SAM ..... S-adenosylmethionine
SDS ..... Sodium dodecyl sulphate
SM ..... Suspension Medium
T ..... Thymine
TAE ..... Tris/Acetate/EDTA (buffer)
Taq ..... *Thermus aquaticus* DNA (polymerase)
TBE ..... Tris/Orthoboric Acid/EDTA (buffer)
TBS ..... Tris-buffered saline
TCA ..... Trichloroacetic acid
TE ..... Tris-EDTA (buffer)
TEMED ..... *N*,*N*,*N*',*N*'-tetramethyl-ethylenediamine
Tris ..... tris(hydroxymethyl)aminomethane
Tris-Cl ..... Tris-hydrochloride
U ..... Units; Uracil or Uridine
UV ..... Ultraviolet
V ..... Volts
Vh ..... Volt hours
W ..... Watts
Xgal 5-bromo-4-chloro-3-inodyl-B-D-galactoside
Z-DNA ..... Z-form (DNA in a left-handed superhelical twist)

Declaration of Intent

In accordance with the University of Edinburgh requirements for submission of a thesis for the degree of Doctor of Philosophy (PhD), I hereby declare that:

(a) The thesis presented here is my own composition.

(b) The work contained throughout is my own, unless stated otherwise. Any information gained from the work of others has been clearly cited and / or referenced throughout the text.

(c) This work has not been submitted for any other degree or professional qualification except as specified.

Signed: Ms. Carmel M. Reilly

Acknowledgements

This work has been carried out under the supervision of Dr. Richard Meehan in the Dept. of Biomedical Sciences, George Square, Edinburgh. Funding was provided by an Edinburgh University Chrichton Scholarship, supplied by the Faculty of Medicine. Other funding was received from the Genes & Development IDG, University of Edinburgh. Dr. Richard Meehan is currently funded by Wellcome Trust and CRC grants.

I would like to thank Richard for all that wisdom, guidance, fast-talking and obscure jokes (I only just got one you told me in '99 the other day), and for remaining so calm over this year when I've been doing my nut. You've been really great to me, thankyou for everything. Thanks also to Jim & Sari and to Irina and Max as work/lab colleagues and for help over the years.

To all the work colleagues & friends I've made- to those who have helped scientifically and to those that have led me astray over the early years of PhD mayhem. Sometimes they are the same person! Big thankyous to Colin, for all of those pub discussions on beer mats about the β-globin gene (!) and for the collaborative work that ensued. Thanks to the lab gang, and to the gang who live outside pipette land in the real world.

I would also like to thank my parents and 'Little C', Catherine for being so supportive over these years, for always cheering me up, putting things in perspective and never taking things too seriously. You really are the best.

And Finally, to end the oscar speech, thanks to those who helped with last minute thesis editing. Sue & Brox for reading...eh 10 pages!, Donncha...thankyou for the music! keeping me sane and driving me insane, sometimes simultaneously. Finally, to Venkat who has helped me more than anyone else over these last few months. Thankyou for all the help & support over these long months in the search for freedom - for always asking was there anything you could do to help, whatever time of day or night it was, and never complaining when I said yes! All my love and thanks.

Abstract

DNA has been shown to exist in a variety of different structural forms. X-ray crystallography has revealed the existence of three distinct tertiary structures, namely A-, B-and Z-DNA. B-DNA is well established as being the predominant form of DNA in the living cell. The biological existence and functional relevance of both A- and Z-DNA is less clear. In addition to these defined structures, DNA can also be induced to bend or kink by a myriad of factors, both intrinsic and extrinsic to the DNA sequence. By default, biological cues must exist to prompt transitions between these forms should they occur *in vivo*, the nature of which, whether physiological, cell-cycle or developmentally based is unknown. It is known that CpG methylation can promote transitions between the A-, B- and Z- DNA isoforms. DNA methylation has also been shown to have effects on both secondary and tertiary DNA structure. These alterations can affect the positioning of the primary DNA packaging proteins *in vitro* and potentially, higher order chromatin structure. A variety of methyl-CpG binding proteins (MeCPs) have now been identified. These in turn interact with a variety of chromatin associated proteins capable of interpreting and reacting to the information provided at methylated CpG sites leading to subsequent changes in gene expression.

Here, I have investigated the structural effects of CpG methylation by a novel method. I have developed an assay that is able to detect CpG methylation induced conformational changes in symmetrically methylated double-stranded DNA. The assay is based on the enzymatic properties of Benzonase, a nuclease that preferentially digests A-form DNA *in vitro.* In this study, a range of DNA gene fragments were used as Benzonase digestion substrates to look for potential A-form regions in CpG methylated DNA. The results suggest that in certain sequence contexts, CpG methylation promotes the formation of A-DNA. These regions of potential A-form DNA are also coincident with sequence motifs that have crystallised as A-form DNA.

In order to test the biological relevance of these results, the affinity of specially designed methylated and unmethylated substrates for a variety of methyl-binding proteins (MeCP2-MBD & xMBD1) was investigated. The results of these experiments are discussed in the text.

From these results, I hypothesise that above the layer of information provided by the DNA sequence, is another layer of coded information provided by DNA methylation. This information is represented by 3-dimensional alterations at both secondary and tertiary structural levels of the DNA helix. These alterations may be in turn, recognised by an array of MeCPs, which in turn interact with proteins which can cause an array of modifications to the nucleosomal histone tails. This tailored display on the histone tails constitutes another set of coded information to influence higher order chromatin proteins, which will operate co-ordinately to effect subsequent localised and/or cellular functional consequences.

Chapter 1: Introduction

1.1 The Significance of CpG Methylation in the cell: An Overview

DNA contains the central core of genetic information of the cell (Avery, 1944). The order of the four bases: adenine, cytosine, guanine and thymine dictates the organisation of a gene (Watson & Crick, 1953). However, not all genes are required to be actively expressed all the time. This would be immensely wasteful and probably lethal. Sometimes, even the two alleles of a diploid cell are used differently, an observation first made by the pioneering work of Barbara McClintock while working on maize kernel colour in the 1950s (Mc Clintock, 1958; Fedoroff, 2002). Throughout development, there are dramatic changes in the pattern of gene expression, which are stably inherited through cell division. As these patterns of expression are reversible, they cannot be explained by DNA sequence variation alone. Consequently, there must be a level of information apart from the primary DNA sequence that specifies the selective use of genetic information during development. The variation of gene expression that is not due to changes in the DNA sequence, is the subject of epigenetics, a term introduced in the 1950s by Conrad Waddington (Waddington, 1957). A modern definition of an epigenetic phenomenon is "a mitotically and/or meiotically heritable change in gene function that cannot be explained by changes in DNA sequence" (Wu & Morris, 2001). The mediators of epigenetic phenomena include, (1) the methylation of DNA at cytosine residues, (2) the acetylation and methylation of histone proteins, (3) the modification and assembly of regulatory protein complexes on DNA and (4) at the post-transcriptional level, a role for small non-coding RNAs These modifications are functionally linked and reversible (reviews: Urnov & Wolffe, 2001; Nakao, 2001).

CpG methylation involves the addition of a 5'-methyl-group to the DNA base cytosine. Figure 1.1 illustrates this process. The modification of CpG methylation is usually associated with gene repression, a link that was first proposed by both Riggs, (1975) and Holliday & Pugh, (1975). The methyl groups are transferred to DNA by DNA methyltransferases (DNMTs) and associated proteins (review: Bestor, 2000; Cheng & Roberts, 2001; Aoki *et al.*, 2001). These may be removed by specific

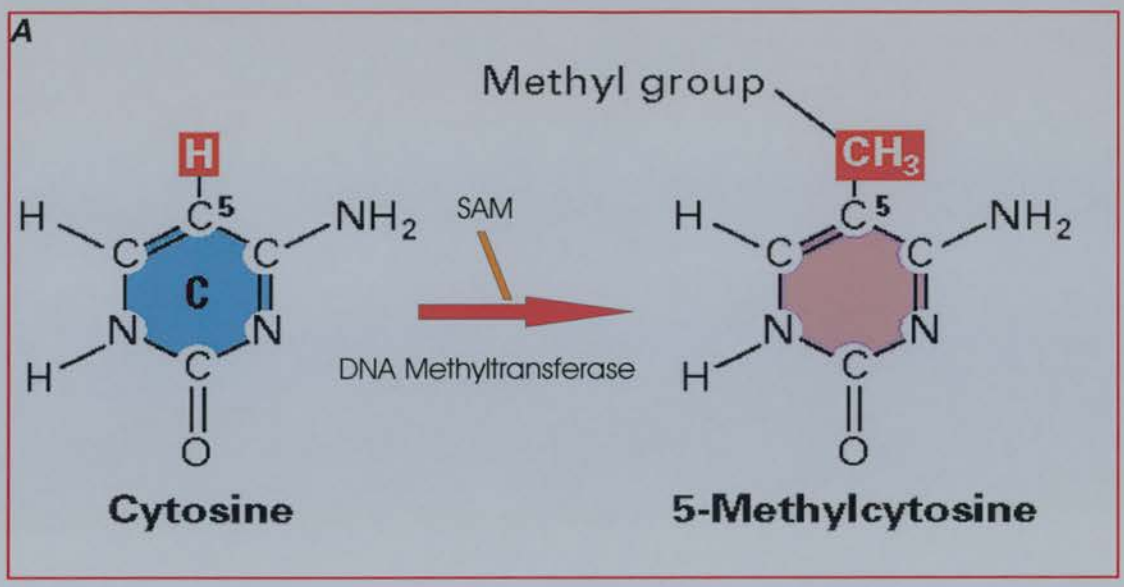
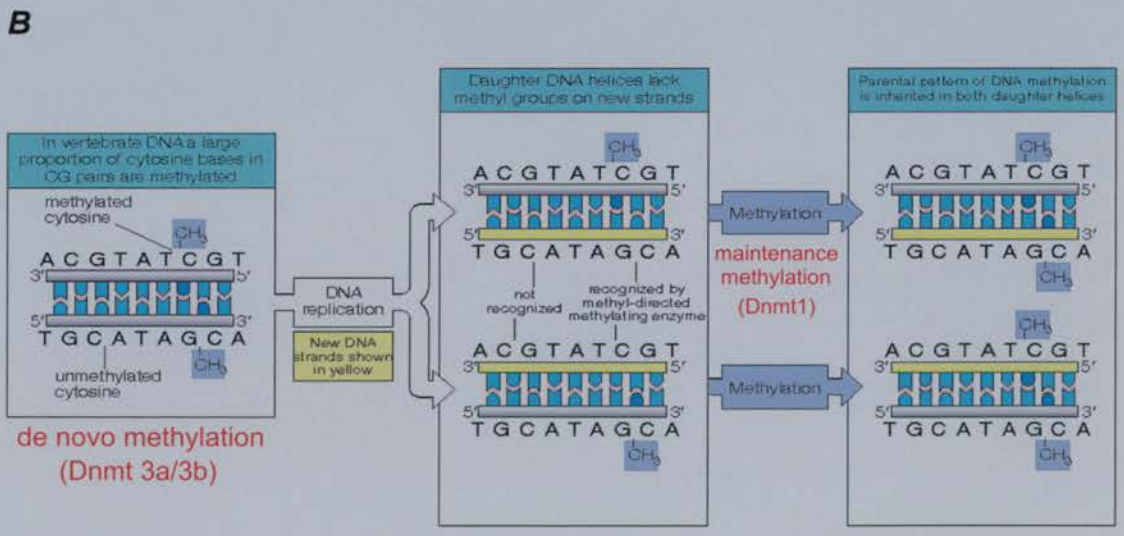


Figure 1.1 *A*, The conversion of cytosine to 5-methylcytosine. A methyl group is transferred from SAM, S-adenosylmethione to the 5' carbon atom of cytosine, by the action of DNA methyltransferases. ***B*, The passage of the replication fork and the establishment of methylation**. After DNA replication DNA strands are rendered hemi-methylated. The hemi-methylated strands are recognised by specific methyltransferases that can use the methylated strand as a template to reciprocally methylate the other strand, thus restoring complete methylation post DNA replication.(Adapted from Lewin, Genes VII).



enzymes, demethylases, but may also be removed by a passive mechanism of post-replication 'phasing out' in the absence of DNMTs (review: Hsieh, 2000; Zhu *et al.*, 2000; review: Kress *et al.*, 2001). CpG methylation can affect gene expression by both direct occlusion of protein binding sites by the methyl-groups and indirect occlusion by proteins binding to methyl-groups (review: Urnov & Wolffe, 2001). Cells can utilise CpG methylation intrinsically to alter both genomic expression levels and also to target extrinsic foreign DNA, such as bacteria, viruses, transposons and or plasmid invaders (Bestor, 1990, 2000, Yu *et al.*, 2001). Indeed, silencing of transgenes and transfection vectors by CpG methylation is a common problem in gene therapy (reviews: Pannell *et al.*, 2001; Morel *et al.*, 2000).

The histone proteins are central to the regulation of CpG methylation processes. They are the proteins that wrap, fold and compact DNA into the cell. In eukaryotes, a central core of eight histone proteins (two copies each of histones H2A, H2B, H3 and H4) wrap 146 bp of DNA through electrostatic interactions in 1.65 left-handed superhelical turns to form the nucleosome particle (Kornberg, 1974, 1977). The linker DNA between nucleosome particles is occupied by histone H1. The combination of nucleosome and histone H1 forms the chromatosome, a basic building block of chromatin (Simpson, 1978). DNA is successively folded through increasing compaction to eventually form the metaphase chromosome. Figure 1.2 illustrates this process. It is on this template that transcription, translation and the majority of genetic control mechanisms must operate. The histones have long basic *N*- terminal tails that extend from the nucleosome core. These tails can contact DNA, other histones and / or other proteins. They can also be modified at specific amino acids to exert different effects on the regulation of gene expression (reviews: Brownell *et al.*, 1996; Csordas, 1990). The now classical association is that CpG methylation and histone deacetylation are associated with gene inactivity and *vice versa*, demethylation and histone acetylation are associated with gene activity (Nan *et al.*, 1998; Jones *et al.*, 1998).

In addition, gene activity can be modified by post-transcriptional RNA-based mechanisms. These mechanisms involve proteins and small RNA molecules that are often coded for in the intergenic regions of DNA (Mattick, 2001). Non-coding RNAs may function as signalling or guide molecules to integrate activity at a locus with

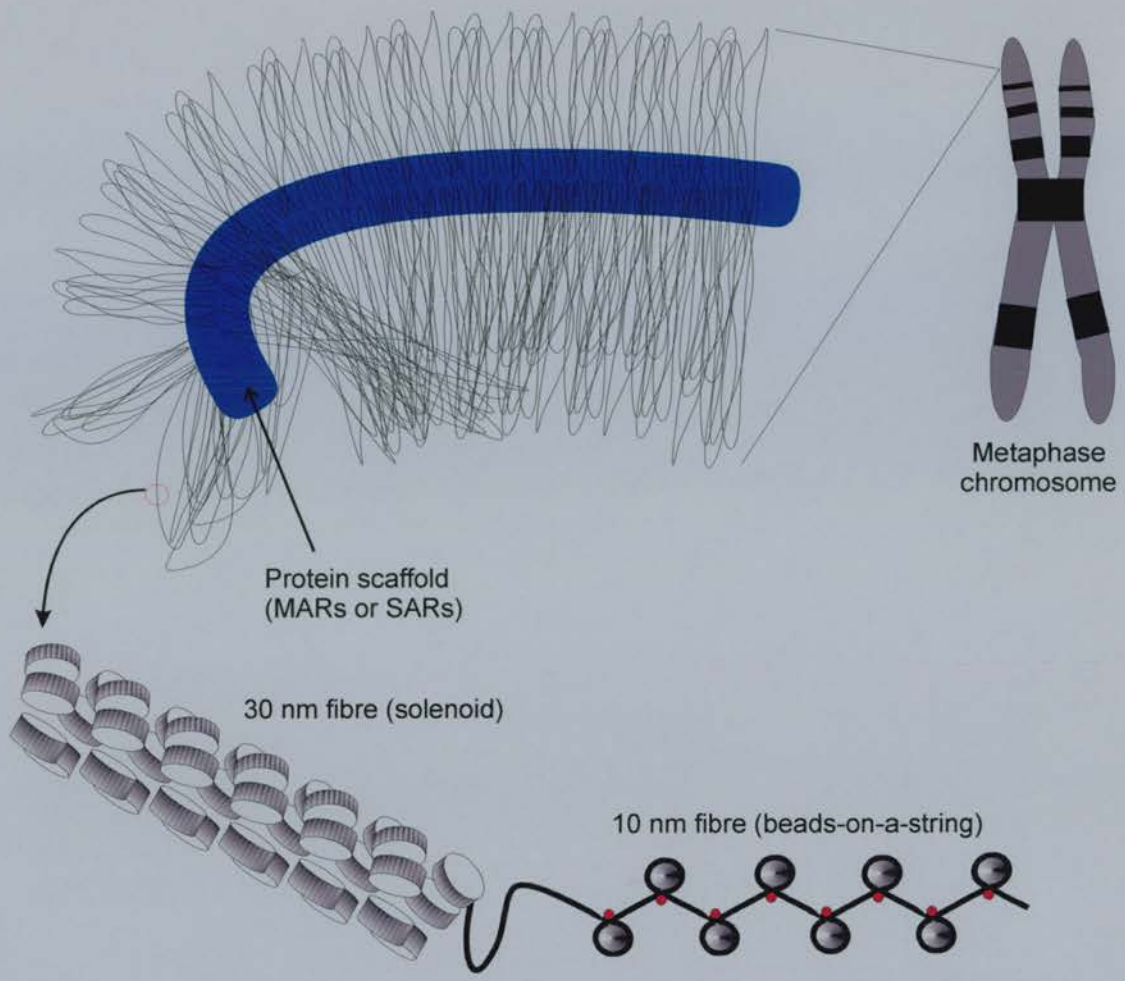


Figure 1.2 Higher order levels of chromatin compaction. Chromatin is folded successively through the 10 and 30 nm fibres and eventually into looped segments anchored on a protein scaffold. Further compaction of this component produces the metaphase chromosome. (Image from Dr. S. Venkataraman, PhD. thesis, 2001).

other parts of the network. These affect transcription, splicing, RNA processing, mRNA translation and chromatin. The RNA-directed regulation of chromatin is homology driven, involving RNA-DNA, RNA-RNA and RNA-protein interactions, in addition to secondary and tertiary structures (review: Matzke *et al.*, 2001). There is evidence to suggest that these RNA based mechanisms may interact with other epigenetic modifications mentioned above at the level of DNA and chromatin, leading to elaborate but elegant methods of gene control (Mette *et al.*, 2000; Santoro *et al.*, 2001; Grishok *et al.*, 2001). Figure 1.3 illustrates the potential complexity of this organised network.

Several epigenetic phenomena in a diverse range of organisms appear to violate Mendelian laws by using some or all of the above epigenetic modifications to alter gene expression. In genomic imprinting, paternally and maternally derived alleles of the same gene are expressed differently (review: Ferguson-Smith, 2001). Another example is 'X-inactivation', a mechanism of dosage compensation to ensure that only one X chromosome is active in the female mammal (review: Park & Kuroda, 2001). Other epigenetic phenomena include 'position effect variegation' (PEV), the silencing of a gene due to its chromatic environment/location on the chromosome (Kellum & Schedl, 1991) and silencing of the mating-type loci in yeast (review: Lustig, 1998).

CpG methylation also plays a distinctive role in a variety of normal and abnormal developmental phenomena. Programmed DNA methylation changes have been linked to early development. In mammals, dramatic remodelling of methylation patterns has been reported for both sperm and oocytes prior to fertilisation (review: Reik *et al.*, 2001). Post-fertilisation, and in early development, methylation levels change less dramatically, a key role of which may be to enhance the expression of specific genes while repressing those that are not urgently required (Reviews: Reik *et al.*, 2001; Meehan & Stancheva, 2001). Evidence for selective expression of genes in development comes from work by Stancheva & Meehan, (2000) who showed that transient depletion of the maternal DNA methyltransferase in Xenopus (*xDnmt1*) leads to premature gene activation of a number of genes. In terms of abnormal development, the processes of apoptosis and cancer are both strongly affected by CpG methylation (Stancheva *et al.*, 2001; Jackson-Grusby, 2001). Hypermethylation

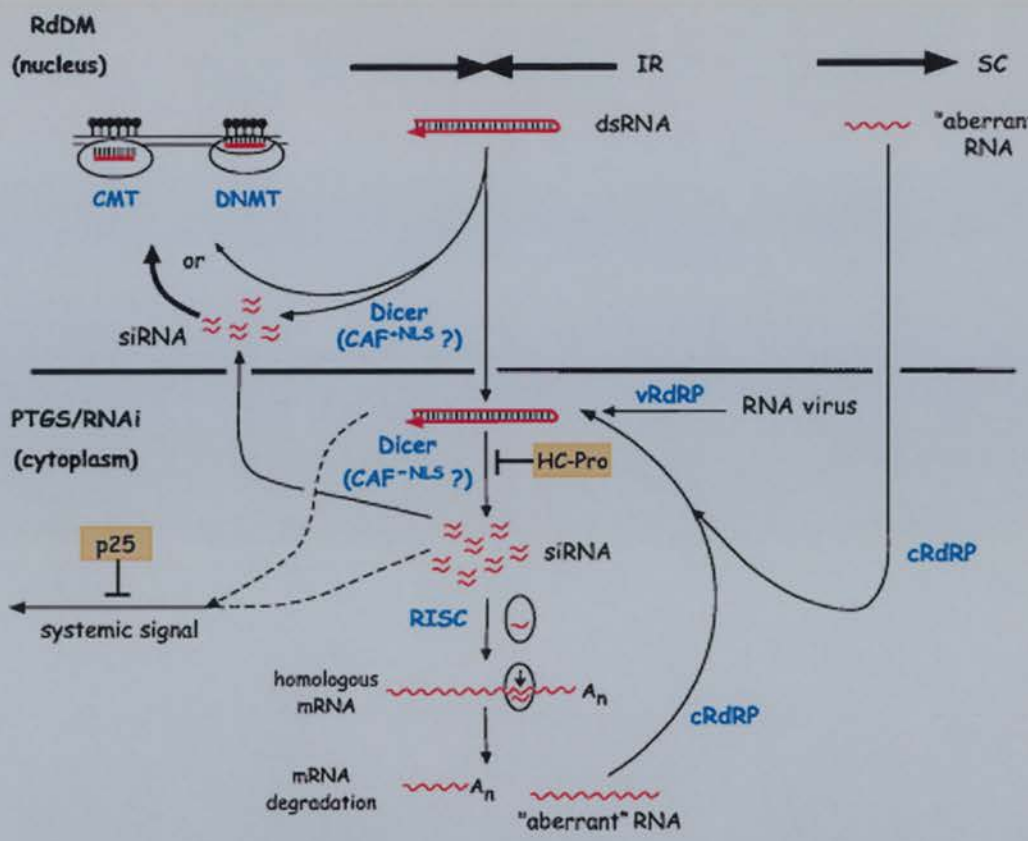


Figure 1.3 A model for RNA mediated silencing in plants. RNA directed DNA methylation (RdDM) and posttranscriptional gene silencing/RNA interferance (PTGS/RNAi) are both triggered by double-stranded RNAs (dsRNAs) that are cleaved by RNAse type III enzymes (eg., Dicer, CAF) into small interfering RNAs (siRNAs), probably in both the nucleus (+ nuclear localisation signal, NLS) and the cytoplasm (-NLS). In the cytoplasm the siRNAs serve as guides for the endonucleolytic cleavage of homologous mRNA degradation in association with the RNA induced silencing complex (RISC). In the nucleus, short RNAs possibly guide methylation of homologous DNA (thick arrow), although dsRNA may trigger RdDM directly. RNA triggering of RdDM may either interact with the Chromomethylase chromodomain (CMT) guiding it to the homologous DNA sequence, or, unwound dsRNA may base-pair with homologous single-stranded DNA, producing an RNA-DNA duplex and single-stranded DNA bulge, an unusual structure that might attract a de novo DNA methyltransferase (DNMT). DNA methylation is represented as filled circles. dsRNAs can be made by transcribing through inverted repeats (IR), or by the activity of cellular RNA-dependent RNA polymerase (cRdRP) acting on 'aberrant' RNA templates synthesised from single-copy (SC) genes in the nucleus or generated in the cytoplasm by RISC cleavage of mRNA. Replicating RNA viruses produce dsRNA by means of a viral RdRP (vRdR). Plant viral suppressors of PTGS, block accumulation of siRNAa (HC-Pro) and systemic silencing (p25). The exact nature of the systemic silencing signal is unknown, but it involves either dsRNA or a special class of siRNAs (dotted lines). (Reproduced and modified from Matzke *et al.*, 2001).

of the usually unmethylated CpG islands associated with gene regulatory regions, renders these genes inactive. This is a common feature in transformed cells. This inactivation can affect virtually all of the pathways in the cellular network. For example, DNA repair (*hMLH1, BRCA1, MGMT*), the cell cycle ($p16^{INK4a}$ $p14^{ARF}$, $p15^{INK4b}$), and apoptosis (*DAPK, APAF-1),* (review: Esteller & Herman, 2002). Methylation also has an adverse effect on somatic cloning technology. A large number of failed cloning attempts have been linked to aberrant methylation reprogramming in the cloned embryos (Dean, 2001; Bourc'his *et al*, 2001).

The mechanism by which DNA methylation exerts these effects is complex. CpG methylation can initially alter DNA structure at both secondary and tertiary levels (Foloppe *et al.*, 1999). These alterations may produce distinct structural patterns that are recognised by a range of methyl-CpG binding proteins (MeCPs). Indeed this has already been speculated on in the context of the rotational positioning of MeCPs on methylated DNA (Wade & Wolffe, 2001), but not in terms of the underlying methylated DNA structure, an aspect that is addressed in this thesis. MeCPs are associated with larger chromatin modifying and remodelling complexes, which help regulate the repression mechanism (review: Ballestar & Wolffe, 2001). At different genes, at different time-points, and in different parts of cell, the exact components of the methyl-binding machinery may alter. This would subsequently alter higher order chromatin structure and affect the formation of inactive chromatin domains (review: Wade, 2001).

Thus, a picture emerges where CpG methylation is involved in many cases as a modification tool for a wide range of epigenetic phenomena. The role of CpG methylation ranges from gene-specific repression mechanisms to genome-wide transcriptional changes in early development. It can also influence a number of cellular processes, a role that when misregulated, can have a catastrophic effect on the cell, leading to apoptosis or worse, the development of transformed and cancerous cells. CpG methylation is regulated by an ever-increasing number of components, at the level of DNA and RNA structure, in addition to a myriad of DNA and chromatin associated regulatory proteins. The result is a complex array of weaving networks that we are only beginning to understand.

So why do we, the mammals possess this unique tool? One speculation is that

the additional layer of control that methylation provides, is in effect our fifth element. It is an evolutionary advantageous tool, adding plasticity to the genome and promoting adaptability. It may have been a key factor in the radiation of the mammals and the subsequent emergence of humans.

1.2 The Effects of CpG Methylation on DNA Bending

In vertebrates, CpG methylation has been harnessed by the cell as a mechanism to control gene expression. A great deal of information has now been gathered in terms of (1) what effects methylation has on DNA structure and (2) what protein(s) and related complexes mediate the effects of methylation of DNA, to regulate transcriptional repression (reviews: Bestor, 2000; Bird, 2002).

In this section, I will describe a variety of different observations on DNA structure. The picture that emerges is one of a very dynamic molecule, capable of many structural alterations, some of which may depend on its methylation status. In some cases, naturally occurring alterations in structure have been used to influence gene expression or to allow protein recognition (Wu & Crothers, 1984). It is important to recognise the fact that whether a gene is expressed or not, depends not only on the order of the genetic code, but also (1) on the propensity of these sequences to alter from the 'idealised' Watson-Crick B-form structure of DNA, (2) how the DNA is packaged into nucleosomes and subsequently, higher-order chromatin (3) where in the cell nucleus the gene is located and (4) the stage of development. Of main concern here is the idea that methylation can alter DNA structure. What these alterations are and what effects they may have on gene expression and / or protein recognition are of direct relevance to this project.

With the publication of the structure of DNA by in 1953 (Watson & Crick, 1953), a blueprint for life was immediately recognisable from the periodicity of the structure (a typical B-form molecule) and the organisation of the bases. The same year Rosalind Franklin published her observations on DNA structure. Her original crystal structures demonstrated an A-form of DNA, the significance of which is investigated here (Frankin & Gosling, 1953). Ironically enough, many years later Crick published the article 'Is DNA really a double helix?' (Crick *et al.*, 1979), fuelled by earlier speculations on the true nature of the structure of DNA (Crick,

1970). An alteration of DNA secondary structure in three-dimensional space is ultimately due to the non-uniformity of the four bases (A, T, C and G). Such alterations range from small kinks or bends, to larger hairpin-loops and cruciform structures. These structures can form in palindromic DNA and are known from mainly bacterial studies, to influence gene expression (Yang *et al.*, 1996).

A collection of studies have focused on methylation of DNA and subsequent analysis of DNA structure, but these have focused on methylation of adenine, A or guanine, G residues at various methylatable sites. These have been considered more important to bacterial genetics, specifically *Escherichia. coli*, and the Restriction-Modification (RM) system (review: Low *et al.,* 2001). Recent evidence now suggests that these modifications may independently influence gene expression in eukaryotes (Lorincz *et al.*, 2001). As we are primarily concerned with CpG methylation and the biological consequences thereof, I am limiting the following to a discussion of predominantly CpG methylation and its effects on DNA structure.

The first direct analyses on CpG methylated DNA structure, were performed in the early 1980's, utilising Deoxyribonuclease I (DNaseI) activity as a probe for DNA structure. The activity of DNase I has been exploited in a number of ways, predominantly for its use in the limited digestion of chromatin. DNase I typically attacks unprotected DNA first, and is known to attack DNA in the minor groove (Drew & Travers, 1984). An early study utilised DNaseI digestion of DNA to investigate potential differences between methylated and unmethylated DNA structure (Fox, 1986). The 5-position of cytosine lies in the major groove; therefore changes in cleavage rates as detected by DNaseI activity can be attributed to DNA structural changes, rather than due to effects of the enzyme itself. The small sequence GCGC was methylated at its central cytosines using *Hha*I methylase. Altered cleavage of this sequence was observed when methylated, the effects being small but consistent changes in local DNA structure. In all cases, 5meC enhanced cleavage of the bond (GpC) on the 5' side of the methylcytosine probably due to the fact that C5 methylation alters the precise orientation of the phosphodiester bonds (Fox, 1986).

Evidence that methylation affects DNA curvature has also been supplied by Diekmann (1987). Different palindromic sequences were ligated to form multimers.

These were differentially methylated and their migration analysed in 10% polyacrylamide gels. The migration differences of these oligos were interpreted in terms of the extent of curvature of the helical axis. It was found that the double-stranded oligomer dCGGAATTCCG is considerably curved and when methylated at the central dAs or dCs, shows an even more pronounced curvature. Multimers of dCGGGATCCCG are straight. Yet when the central dAs or dCs are methylated, a small curvature is induced (Diekmann, 1987). This could be considered evidence enough, at least *in vitro*, that methylation (whether at dAs or dCs) has a general effect on the curvature of a piece of DNA. However, three aspects should be borne in mind with respect to many such observations. **(1)** The degree of curvature seen when sequences are methylated is dependent on the degree of curvature seen in the unmethylated oligonucleotide and thus directly on the sequence of the oligomer. **(2)** The number of base pairs involved, whether they have the potential to form hairpin or cruciform structures and what proportion of the helical axis is being represented are pertinent questions that were beyond the scope of this analysis. **(3)** The above study also demonstrated that multimers of the sequence d(CCATCGATGG), when methylated at the central most dAs on both strands, has no effect on electrophoretic mobility. This latter point, while demonstrating that changes in gel mobilities are not due to an increase in molecular weight, nevertheless dissuades from an argument for methylation induced curvature acting as a recognition element for DNA-protein interactions.

To differentiate between the effects of N6-adenine and C5-cytosine methylation on the kinetics of cruciform extrusion, and thus on helix stability, Murchie *et al.*, (1989) differentially methylated a selection of inverted repeats containing central restriction sites by using bacterial methylases. The rate constants for cruciform extrusion at 37°C were measured. Methylation was found to have a pronounced affect on the rate of salt-dependent cruciform extrusion. To form cruciform structures from normally base-paired and unmethylated templates, it is necessary to have some initial opening of the helix. A mechanism of salt-dependent (S-type) cruciform extrusion has been proposed (Murchie & Lillie, 1987). It would be expected that helix stability, being dependent on base-sequence, would have a large influence on the rates at which cruciform structures are formed. From this

study it was found that A-methylation weakens the double helix, while C-methylation results in stabilisation (Murchie *et al*, 1989). This is in accordance with other studies of base methylation (Ehrlich *et al.*, 1975; Collins & Myers, 1987; Yamaki *et al.*, 1988). All generally conclude that N6-methylation of adenine directly affects base pairing by destabilising hydrogen bonding interactions with thymine. Methylation at C residues is not at a position that would produce the same effect. The strength of this study is that it confirms that methylation at A and C residues have different structural effects, with C5 methylation capable of changing local helix stability constants by about 10% in supercoiled substrates (Murchie *et al.*, 1989). The corollary of this being that demethylation of meCpG-rich regions could facilitate local unwinding *in vivo*, and thus have a direct influence on gene expression.

Sequences of the form (dA)n.(dT)n (A-tracts), both natural and synthetically derived, are known to exhibit pronounced degrees of stable curvature of the helix axis. This is dependent on their being in phase with the helical repeat (Koo & Crothers, 1987). Why exactly this should be so is unclear. Hagerman, (1990a) investigated further what sequences potentiated DNA curvature (review, Hagerman 1990 b). The effects of the 5-methyl group on DNA curvature were examined in a series of synthetic molecules in which the position of the 5-methyl group was varied for both A- and I-tract sequence motifs. Pyrimidine 5-methyl groups were found to have a significant effect on the degree of curvature. This was also dependent on their locations within the homo-oligomeric sequences. These effects were observed for both $(dA)n.(dT/dU)_n$ and $(dI)_n.(dC/d5meC)_n$ sequence motifs. Some previous experimental and computational studies had argued against a model of 5meC induced bending (Koo & Crothers, 1987).

However, from this and other more recent work, it is now reasonable to accept that methyl groups do significantly alter the degree of curvature in a position dependent fashion. That is, in a follow up study to the above, Hodges-Garcia & Hagerman (1992, 1995) investigated the effects of cytosine methylation specifically, on the structure of duplex DNA. This work demonstrated that the effects of cytosine methylation on A-tract DNA were mediated by the methyl group, and were not a consequence of being inserted within an already stably curved A-tract. Electrophoretic mobility analysis led to the observation that methylation of cytosine

residues up to three base pairs away from a $(dA)_5.(dT)_5$ tract (A-tract), still altered the net curvature of DNA. This depended on both the sequence of the non-A-tract and the positions of the methylated residues. In addition, the decamer CAAAAAGme^{5}CCG is characterised by a bend much more pronounced than the unmethylated decamer, but for the sequence CAAAAAme^{5}CGCG, methylation only slightly modifies the DNA curvature (Hodges-Garcia & Hagerman, 1992). Further to this, NMR analysis has revealed that the magnitude of the effect of CpG methylation is dependent on the sequence context. ^{1}H-NMR and ^{31}P-NMR experiments have been used to compare the octamers $d(CATCGATG)_2$ and $d(CTTCGAAG)_2$. CpG methylation of the former led to a larger conformational variation than the latter (Lefebvre *et al.*, 1995a & 1995b).

In conclusion, these results add weight to the main hypothesis explored in this thesis; that the biological consequences of methylation are mediated in part by an altered DNA structure, as well as via protein-DNA interactions. A set of 'rules'- analogous to the rules that govern nucleotide coding sequences and amino acid production, may govern whether a CpG methylated site will have a positive or negative regulatory role. This would be dependent on immediate sequence context, position within a gene, regulatory region or elsewhere, and then on whether the proteins required are available to bind. Indeed, the number of methylated sites available, and thus the degree of curvature exhibited, may act as the combined biological cue for protein binding and subsequent transcriptional effects.

In relation to such observations, what is lacking from the above studies is a direct functional relevance of methylation-induced structural changes on alterations in gene-expression *in vivo*. Sequence-specific methylation of eukaryotic RNA polymerase II- and III- transcribed promoters can cause their inactivation. The precise reason for this was investigated in terms of the structural effects that methylation causes in such promoters (Muiznieks & Doerfler, 1994). Three DNA elements were chosen. The late E2A promoter of adenovirus type 2 (Ad2) DNA, (Ad2E2AL), the VAI (virus-associated) RNA gene of Ad2 DNA, and an Alu element associated with the human angiogenin gene. These elements were specifically *in vitro* methylated with either the *Hpa*II, *FnuD*II or the CpG DNA methyltransferase (M-SssI) from *Spiroplasma* species. The effects of methylation by these enzymes were analysed in

terms of their electrophoretic mobilities. The E2AL promoter fragment was shown to have a sequence-dependent bend consistently at a TACA sequence. In addition, abnormal migration of sub-fragments of the E2AL promoter confirmed the location of majorly bent DNA to be at the so-called TACA motif. Quantification of the curve mathamatically, revealed that fluctuations of sub-fragments from the estimated value, suggested an influence from other DNA 'motifs' on DNA curvature. Similar effects of methylation on DNA curvature were seen with the other promoters analysed (Muiznieks & Doerfler, 1994). In all cases, the extent of curvature seen was dependent on the position of methylation in a given methylatable motif and the position of that motif in terms of overall DNA topology. Critically, all of the promoters demonstrated that a few 5-mC residues in 5'-CG-3' dinucleotides can elicit reproducible changes in DNA topology.

In the case of the E2AL promoter, the critical sequence for bending lies 5' adjacent to the TATA protein-binding motif. Both the TATA-binding protein and the transcription factor TFIID bend DNA upon binding. The effect that changes in DNA topology and bending could have on this interaction was investigated by methylation of a single CpG site of the E2AL promoter *in vitro*, which was placed under the control of a CAT reporter construct and transfected into HeLa cells. Methylation of a single 5'-CCGG-3' site reduced CAT activity. Progressive methylation of all 5'-CG-3' sites in the pAd2E2AL-CAT construct, completely abolished CAT expression from this promoter, indicating a decisive role for methylation induced changes influencing gene expression (Muizineiks & Doerfler, 1994). The question of how methylation induced changes in DNA structure directly affect TATA-protein binding at this promoter, were not addressed. However, several important conclusions have resulted from this work;

(1) The available data suggests, that methylation of individual 5'-CG-3' sequences, although capable of inducing minor kinks in DNA, cannot alone be responsible for significant bending or curvature of DNA over long stretches, given that many individual kinks may have opposing effects and cancel each other out. It is possible that long stretches of CG-rich and methylated DNA may significantly alter DNA structure such that gene expression is also significantly altered. However, it is also supposed that methylation-induced alterations in DNA structure at single CpG

sites may have a profound effect on gene expression should these positions hinder protein binding such as at promoters, enhancers and related control regions. Interestingly, this study also showed that 5'C methylation of either 5'-CG-3', 5'-GCGC-3' or 5'-CCGG-3' sequences induces similar DNaseI hypersensitive sites 5' to the methylated C residue. In this vein, a prime example whereby small changes in methylation status have a profound effect is that of methylation induced changes in nucleosome positioning at the chicken adult β-globin gene. Here, methylation at a CpG triplet can shift a specific nucleosome from an otherwise strong positioning site, implying that methylation-induced changes influencing chromatin structure directly (Davey *et al*, 1997).

(2) Methylation induced curvature of 5'-CCGG-3' fragments is also more striking when methylated sites are adjacent to each other and at the centre of the molecule. Perhaps this is significant to a potential cumulative effect of many adjacent methylated CpG dinucleotides.

(3) Although DNA bending effects are clearly seen with methylated 5'-CG-3' sequences, these effects are not apparent when cytosine residues in the sequence 5'-CC(A/T)GG-3' are methylated. This could imply that bending is specific for methylated 5'-CG-3' sequences, however from examples cited earlier and subsequently, this cannot be considered as the only case. eg: methylation of adenine in the sequence 5'-GATC-3' by the dam methyltransferase affects bending very strongly (Muiznieks & Doerfler, 1994).

From the evidence above, it is more than reasonable to conclude that methylated DNA has an altered DNA structure, which in the case of 5'meC is dependent on a number of factors. These may include the sequence context of the methylatable substrate, the length of the substrate involved and GC-richness, among others. Given that we are primarily concerned with methylated d(CG)n sequences, a number of these variables can be stabilised, the effects of which will be discussed later. Structural changes in methylated DNA may be part of the initial mechanism by which its many effects are executed. As mentioned previously, DNA has now been crystallised in A-, B- and Z- forms. Methylation of 5'C residues has also been shown to have a direct effect on whether such alternative forms exist. Thus, crystallography of methylated sequences of varying length and sequence context has shown

interesting results pertaining the general structure of methylated DNA.

1.3 DNA Conformation: A-, B- and Z-DNA

The elucidation of the structure of DNA by Watson and Crick (1953) heralded a new era for the biosciences. It is this structure that is most familiar to us all as the right-handed double helix of B-DNA (figure 1.4), the prevalence and biological significance of which has been firmly established. However, other structures do exist (table 1.1). The priorities here are the increasingly studied forms of A- and Z- form DNA. A-DNA (figure 1.5), is also a right-handed helix, but differs from B-DNA in how tightly its bases are stacked together. As a result, it has the appearance of an overwound B-DNA molecule with a hole stretching along the helical axis. Z-DNA adopts a highly unusual left-handed helical structure which differs considerably from both B- and A- form DNA (figure 1.6).

DNA is in the B-conformation when the counterion is Na^+ and the relative humidity is 92%. It is regarded as the native form. The two plectonemically coiled polynucleotide strands wind about a common axis with a right-handed twist, forming a 20 Å diameter double helix. The base pairs form the core of the helix about which the sugar-phosphate chains are coiled. The ideal B-DNA helix has 10 bp per turn and therefore a helical twist of 36° per bp. Since the bases have a van der Waals thickness of 3.4 Å and are partially stacked on each other, the helix has a pitch (rise per turn) of 34 Å (Voet & Voet, 1990). What is critical in terms of protein recognition of DNA is the shape of the major and minor grooves. As seen from table 1.1, B-DNA maintains a wide and deep major groove, the minor groove being narrow and deep. Many years later, Wing *et al.,* (1980) revealed that the self-complementary dodecamer d(CGCGAATTCGCG) crystallises as B-DNA, but with slightly different parameters to that of the Watson-Crick structure. This raised the point that individual residues can contribute to the overall structure differently, in an independent fashion. B-DNA could be considered the average output from any given randomly generated sequence. When sequences are decidedly non-random in their sequence bias, such as in CG-rich regions of DNA, the structure generated can differ considerably. As will be discussed later, the effect of CpG methylation on these transitions may have a large effect on protein recognition of DNA and subsequently gene expression. At a

	A	B	Z
Helical sense	Right handed	Right handed	Left handed
Diameter	~ 26 Å	~ 20 Å	~ 18 Å
Basepairs per helical turn	11	10	12 (6 dimers)
Helical twist per base pair	33°	36°	60° (per dimer)
Helix pitch (rise per turn)	28 Å	34 Å	45 Å
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C(3')-*endo*	C(2')-*endo*	C(2')-*endo* for pyrimidines; C(3')-*endo* for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines; syn for purines

Table 1.1 Structural features of idealised A-, B- and Z-DNA tertiary structures. Of note is that both A- and B-DNA forms are right-handed helices. Z-DNA is left-handed. The major groove of A-DNA is narrow and deep as opposed to the wide and deep major groove of B-DNA and the almost flat major groove of Z-DNA (adapted from Voet & Voet, 1990).

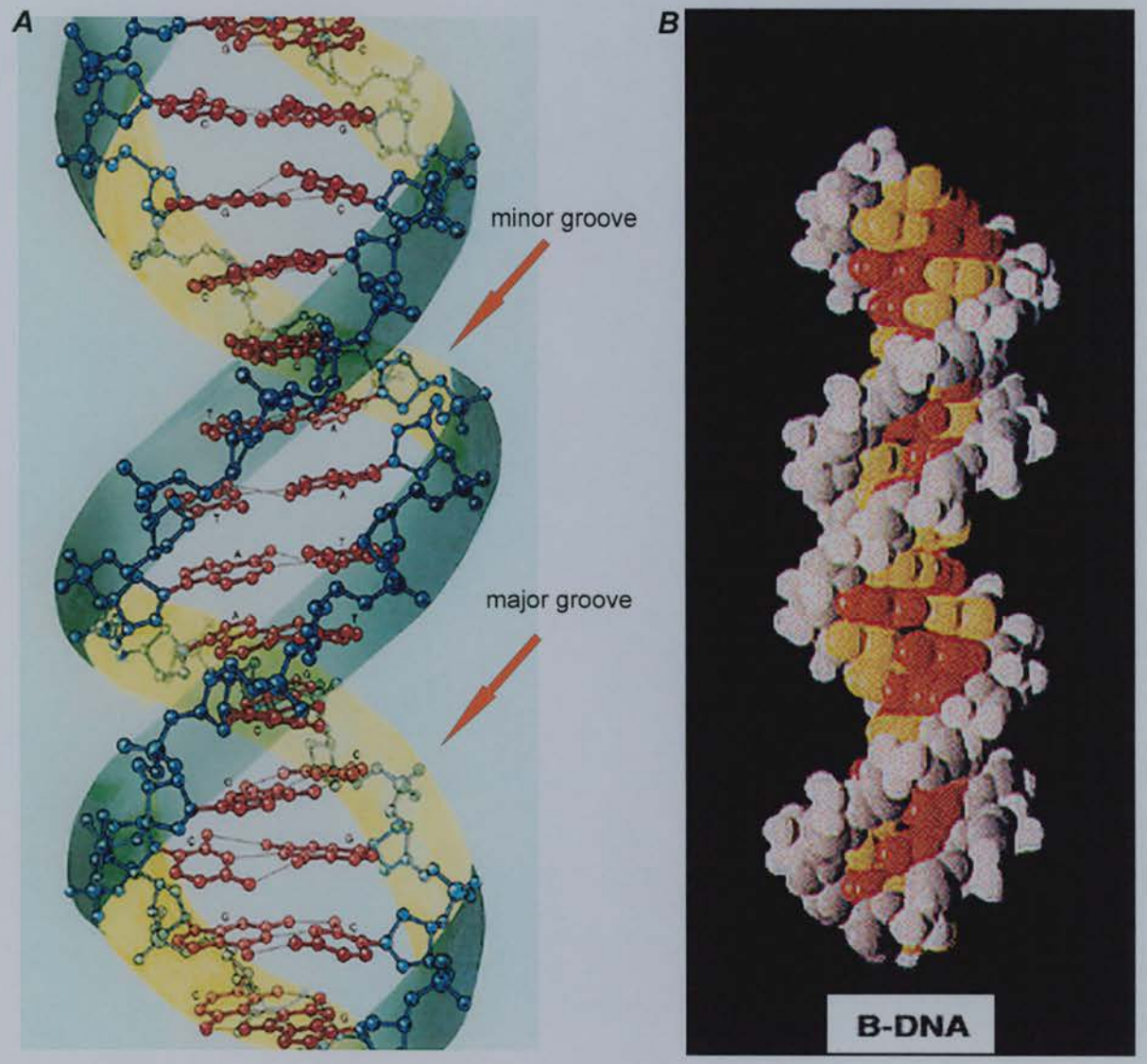


Figure 1.4 The structure of B-DNA. The repeating helix is based on the X-ray structure of the self-complementary dodecamer d(CGCGAATTCGCG) determined by Dickerson & Drew (1970). The view is perpendicular to the helix axis. ***A,*** ball and stick representation of B- DNA. In blue, the sugar-phosphate backbones, winding about the periphery of the molecule. In red, the bases which occupy the core. ***B,*** space-filled model of B-DNA. The sugar-phosphate backbone is white and grey, with the bases occupying the central core of the molecule. As the helix axis passes through the base pairs, the helix has a solid core. (Drawing by Irving Geis, adapted from Voet & Voet, 1990; and http://ndbserver.rutgers.edu/NDB/NDBATLAS).

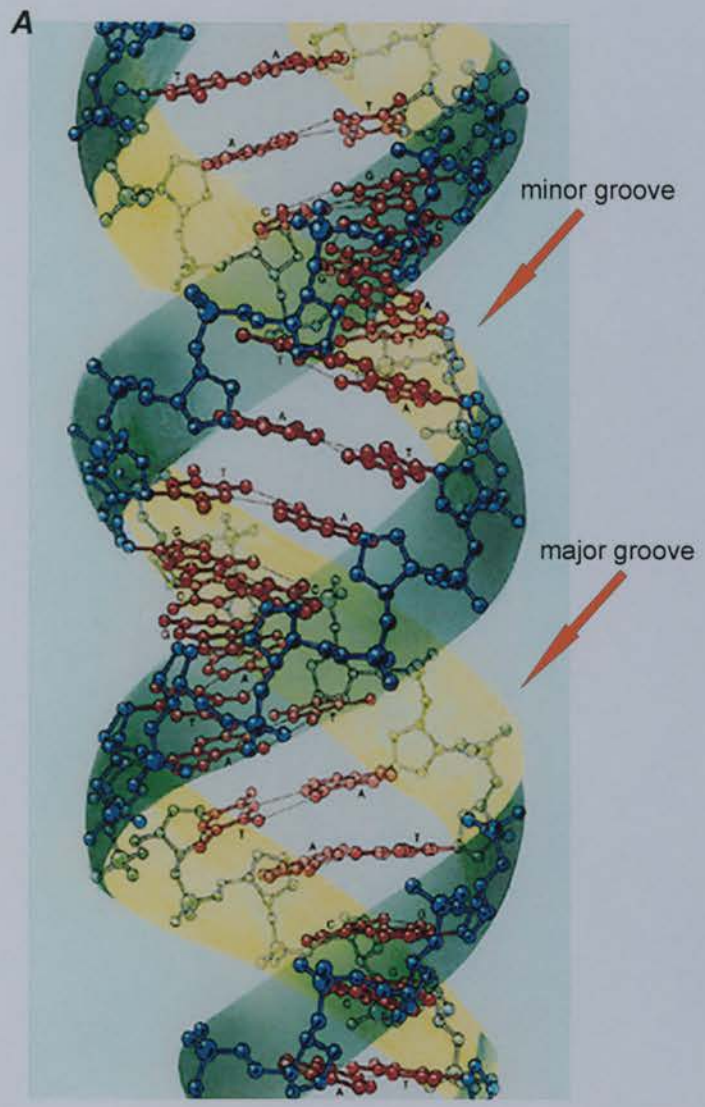


Figure 1.5 The structure of A-DNA. The repeating helix was generated by Dickerson based on the self-complementary octamer d(GGTATACC) as viewed perpendicular to the helix axis. ***A,*** the ball and stick model and ***B,*** the space-filled model. Note that the base pairs are inclined to the helix axis and that the helix has a hollow core. (Drawing by Irving Geis, adapted from Voet & Voet, 1990; and http://ndbserver.rutgers.edu/NDB/NDBATLAS/).

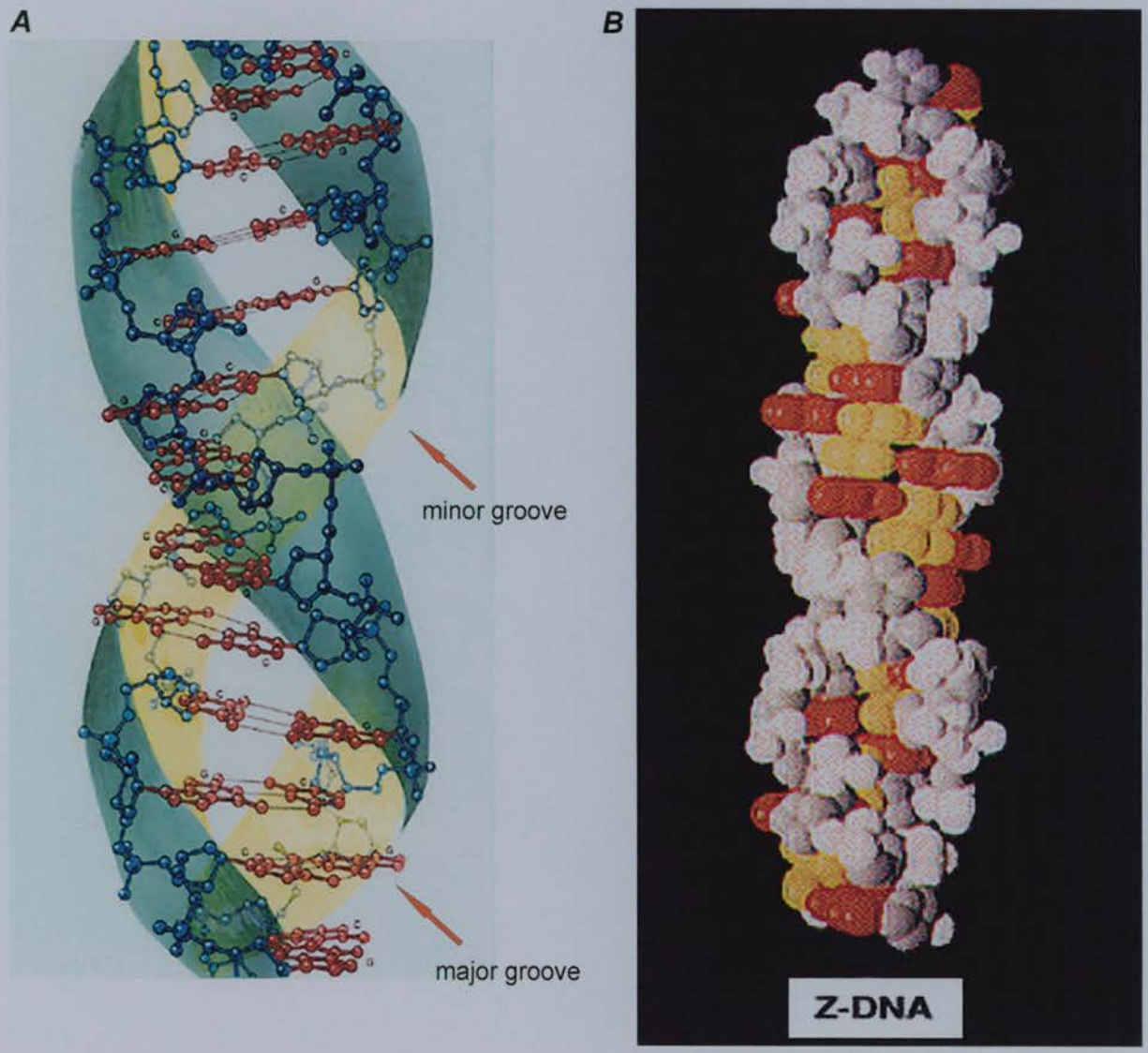


Figure 1.6 The structure of Z-DNA. The repeating helix was generated by Dickerson based on the X-ray structure of the self-complementary hexamer d(CGCGCG) determined by Wang and Rich (1984).The view is perpendicular to the helix. ***A,*** the ball-and-stick model and ***B,*** the space-filled model. Note that the helix is left-handed and that the sugar phosphate chains follow a zig-zag course with alternate ribose residues lie at different radii.(Drawing by Irving Geis, adapted from Voet & Voet, 1990; computer graphic from http://ndbserver.rutgers.edu/NDB/NDBATLAS.).

relative humidity of 75%, a transition from B-DNA to the wider, flatter A-DNA is possible. A-DNA has 11 bp per turn and a pitch of 28 Å, which provides A-DNA with an axial hole. Because its bases are tilted 20 Å with respect to the helical axis, A-DNA possesses a deep major groove and a very shallow minor groove. Typically, self-complementary oligomers of less than 10 bp can be crystallised in the A-conformation (Voet & Voet, 1990). In a functional context, there is some evidence to suggest that A-DNA is formed in the presence of small acid-soluble proteins called (SASPs) in sporulating bacteria such as *Bacillus subtilis.* These proteins are activated after the onset of sporulation and protect the spore from UV damage. SASP protein binding to DNA changes DNA conformation from B to A-form, whose stereochemical constraints inhibits thymine-dimer formation (Mohr *et al.*, 1991). More recent evidence suggests that A-DNA may have a broader functional role in that CpG methylated DNA can assume an A-DNA conformation under physiological conditions (Dickerson & Ng, 2001; Hodges-Garcia & Hagerman, 1995). This could form the structural substrate for proteins known to recognise CpG methylated DNA (this thesis).

Twenty-six years after the discovery of the Watson-Crick structure, Wang *et al.,* (1979) solved the crystal structure of d(CGCGCG) as being a left-handed double-helix which they termed Z-DNA. d(CGCATGCG) also forms a similar structure (Wang *et al.*, 1979). Z-DNA maintains 12 Watson-Crick base pairs per turn and a pitch of 45 Å. As a result, Z-DNA has a deep minor groove and no obvious major groove. The base-pairs are flipped 180° relative to B-DNA, and the repeating unit is that of a dinucleotide pair following a zig-zag pattern, very unlike the other structures (figure 1.6). Alternating purine-pyrimidine steps in high salt conditions, are the most favoured for this conformation. Having established that this structure does exist, the function of its existance has been more elusive. Various interpretations exist. Sera obtained from patients with autoimmune diseases have been shown to produce antibodies specific for Z-DNA (Lafer *et al.,* 1983). In addition, both the puffed regions associated with transcriptionally active polytene chromosomes in *Drosophila melanogaster* (but not in the banded regions normally associated with heterochromatin) (Lancillotti *et al.,* 1987) and the transcriptionally active macronucleus of ciliated protozoa (but not the transcriptionally dormant

micronucleus), show reactivity to Z-DNA antibodies. Some evidence suggests that Z-DNA forms behind a transcribing RNA polymerase, stabilised by negative supercoils produced in transcription (review: Hill, 1991). A great deal more is known about Z- than A-form DNA. For this reason and to highlight the functionality of different conformations of DNA, this will be discussed below in addition to current observations on the A-form structure. Of direct relevance to this thesis is the effect that CpG methylation has on the A-, B- and Z-DNA. This will be discussed below.

1.4 The Effects of CpG methylation on DNA Conformation

1.4.1 Z-DNA

The advent of X-ray crystallographic techniques in the early 1960's (Kendrew *et al.*, 1960) and the subsequent development of triester methods of DNA synthesis (Itakura, *et al.*, 1975) together led to the first direct analysis of highly purified crystals of DNA of predetermined sequence. These studies led to the discovery of left-handed Z-DNA (Wang *et al.*, 1979; Drew *et al.*, 1980). A variety of sequences have now been crystallised in the Z-conformation. These crystallographic forms are summarised in table 1.2.

The B to Z transition is favoured by alternations of purines/pyrimidines. Typically, dC/dG alternating sequences are a favoured substrate. Left handed Z-DNA has been observed by X-ray crystallography and NMR for a variety of synthetic d(CpG) oligos of varying sizes, in addition to poly(dG-dC).poly(dG-dC) and poly(dA-dC).poly(dG-dT) (Wang *et al.*, 1979; Drew *et al.*, 1980; Crawford *et al.*,1980; Arnott *et al.*, 1980; Pohl *et al.*, 1972). However although these studies clearly demonstrated the existence of Z-DNA, problems do exist with their comparison, given the conditions under which the crystals are prepared vary greatly.

The analysis of DNA structure by X-ray crystallography, NMR and other methods, extended into looking at the effect of methylation on DNA conformation. Behe and Felsenfeld (1981) compared various methylated and unmethylated oligonucleotide substrates for their efficiency of the B->Z transition. The synthetic oligonucleotide poly(dGm^{5}dC).poly(dGm^{5}dC) was found to be capable of forming Z-DNA, but at salt concentrations much closer to physiological conditions, unlike its

Sequence	Type of DNA	Reference
A. *Pyrimidine start*		
CGCG	Z	Drew *et al.* (1980)
CGCGCG	Z	Wang *et al.* (1979)
CACGTG	Z	Coll *et al.* (1988)
CGCGTG	Z	Ho *et al.* (1985)
(CACGCG)•(CGCGTG)	Z	Sadasivan & Gauthan (1995)
(CGCACG)•(CGTGTG)	Z	Sadasivan & Gauthan (1995)
CGCATGCG	Z	Fugii *et al.* (1985)
CGCGCGCG	Z	Fugii *et al.* (1985)
CGCICICG	Z	Kumar *et al.* (1992)
CGTACGTACG	Z	Brennan *et al.* (1986)
CGCATATATGCG	B	Yoon *et al.* (1988)
$m^5CGm^5CGm^5CG$	Z	Fugii *et al.* (1982)
$br^5CGbr^5CGbr^5CG$	Z	Chevrier *et al.* (1986)
m^5CGUAm^5CG	Z	Zhou & Ho (1990)
B. *Purine start*		
GTGTACAC	A	Jain *et al.* (1987)
GTACGTAC	A	Takusagawa (1990)
ATGCGCAT	A	Clark *et al.* (1990)
GTACGTAC	A	Langlois *et al.* (1993)
GTGCGCAC	A	Bingman *et al.* (1992a)
GCGTACGTACGC	A	Bingman *et al.* (1992b)
Gm^5CGCGC	A	Moores *et al.* (1995)
Gm^5CGm^5CGC	A	Moores *et al.* (1995)
GCACGCGTGC	A	Ban & Sundaralingam (1996)
GCGCGCGCGC	Z	Ban *et al.* (1996)
$Gm^5CGm^5CGCGCGC$	A	Tippin *et al.* (1997)
$d(GGCGm^5CG)_2$	A	Vargason *et al.* (2000)
$d(GGCGm^5CC)_2$	E	Vargason *et al.* (2000)

Table 1.2 Crystal structures of alternating oligonucleotide sequences. Highlighted in red, m^5C, is 5-methylcytosine; in blue, br^5C, is 5-bromocytosine; in yellow, CGCATATATGCG, containing three consecutive AT base-pairs, crystallises in a right-handed conformation and is the only B-DNA sequence included in this list; in purple, GCGCGCGCGC, which contains four consecutive CG base-pairs, is the only purine start sequence that crystallises in a left-handed conformation. E-DNA is a stable intermediate in the transition of B-DNA to A-DNA (Adapted from Tippin *et al.*, 1997).

unmethylated counterpart. In addition, experiments involving mixed copolymers of dC and m^5dC, indicated that the introduction of methyl groups has a direct effect in lowering the transition point between B and Z-forms. With 30% of the cytosine residues methylated, the B->Z transition occurs at a Mg^{2+} concentration 1/35th of that required for the completely unmethylated polymer (Behe & Felsenfeld, 1981). This was confirmed in similar studies by Takeuchi *et al.*, (1994) and additionally indicates that a small number of methylated sites have the potential to exert long-range structural changes on alternating CG dinucleotides.

Subsequent work by Fujii *et al.*, (1982) demonstrated similar results in that the synthetic oligo $(m^5dC\text{-}dG)_3$ formed Z-DNA by single crystal X-ray diffraction analysis. The Z-structure is similar to that of the unmethylated form, with the methyl group causing slight changes in the twist angle between successive base pairs (Fujii *et al.*, 1982). The above analyses were all carried out on isolated fragments of DNA. Although alternating purine-pyrimidine (dC-dG)n sequences are generally considered the ideal structure for Z-DNA formation, NMR analysis of the decamer $d(br^5CGbr^5CGATbr^5CGbr^5CG)$ and a similar dodecamer, have revealed this too is capable of forming Z-DNA (Feigon, *et al.*, 1985; Hua *et al.*, 1989). Of note here is that despite the presence of the central AT pair disrupting the sequence alternation, this decamer is still capable of forming Z-DNA. This is apparently aided by the bromination of cytosine, a modification very similar to C5 methylation. Unfortunately, this is aided by high salt and methanol conditions, but perhaps if this is also true for C5 methylation *in vivo,* these conditions are compensated for by a stabilisation due to MeCP binding properties (Wade & Wolffe, 2001).

Klysik *et al.*, (1983) used circular dichroism of both restriction fragments and recombinant plasmids to observe the B->Z transition. Unlike the previous examples, ionic strengths higher than physiological levels were required to form Z-DNA, when the methylated Z-forming region is adjacent to typical non-Z forming regions in both fragments and plasmids. Therefore, the effect of B-Z junctions is significant in the promotion of Z-DNA via methylation *in vitro,* even though these levels were not as high as those required to form Z-DNA from non-methylated templates. Z-DNA can also be stabilised by negative supercoiling. In this regard, supercoiled plasmids containing up to $(dC\text{-}dG)_{29}$ methylated sequences were found to significantly reduce

the number of negative supercoils required to stabilise the Z-conformation. Concluded from this work was the prospect that constrainment of DNA supercoiling within nucleosomes of eukaryotic DNA could result in the elimination of a trigger for Z-DNA formation, a trigger that DNA methylation was subsidising (Klysik *et al.*, 1983).

The power of methylation to induce changes in DNA conformation in both a structural and energetic sense was determined by Zacharias *et al.*, (1988) using 2-D gel electrophoresis. Plasmid constructs containing either dC-dG or m5dC-dG inserts were compared in terms of their ΔG^{o} values (free energy for formation) of Z-DNA based on various structural parameters relating to the B->Z transition. Quantitatively, comparisons between energies required to form Z-DNA revealed that the methylated insert significantly aided the transition to Z-DNA. Also qualitatively, a plasmid construct containing two dC-dG blocks separated by 95bp of random sequence underwent independent B->Z transitions whereas the same dC-dG blocks separated by only four base-pairs (GATC), performed the transition to Z-DNA as one unit, which included the 4 bp linker region (Zacharias *et al.*, 1988). This indicates that the extent of methylation and the distance between methylated sites are both important in relation to the formation of new structures. Also, there may not need to be a requirement for successive CpG sites to be methylated for an overall change in DNA structure to occur, and hence the possibility exists that this may have some bearing in terms of methylation-mediated repression of gene expression.

More recently, Zacharias *et al* (1990) have investigated the effect of cytosine methylation on Z-DNA formation *in vivo*. DNA supercoiling has been found to significantly stabilise Z-DNA. Importantly this work looked at supercoil-stabilised *E. coli* DNA as plasmid inserts. Cytosine methylation was found to directly enhance Z-DNA formation of dC-dG segments of varying length. The B->Z transition was promoted by utilising *Hha*I methyltransferase (M.*Hha*I) and for all lengths of inserts methylation enhanced the transition at the superhelical density present in *E. coli.* Overall, this work further indicates that methylation influencing the B->Z transition could constitute a mechanism for the modulation of DNA-protein interactions and gene expression.

Although it is now firmly established that methylation significantly promotes

the transition to Z-DNA, the purpose of this and indeed the general significance of Z-DNA in its own right are questionable. The initial focus of Z-DNA analysis was in terms of the purine/pyrimidine alternations that promote this structure. In the absence of methylation and supercoiling effects, Z-DNA can also be stabilised by polyamines such as spermine, and specific Z-binding proteins (review; Herbert & Rich, 1996).

Efforts to uncover the potential biological role of Z-DNA have focused on uncovering potential Z-DNA binding proteins. The quest for Z-DNA binding proteins has thrown forth a variety of puzzling candidates. The first studies isolated proteins from SV40 DNA that bound to negatively supercoiled, Z-DNA containing sequences (Azorin & Rich, 1985). Since then, other examples come from the isolation of a *Drosophila* Z-DNA binding Topoisomerase II homolog (Glikin *et al.*, 1991; Arndt-Jovin *et al.*, 1993), Z-DNA-binding proteins from bull testis (Gut *et al.*, 1987) and an example of the bovine ocular lens zeta-crystallin protein (Gagna *et al.*, 1998). Apart from these disparate examples, the main focus in recent years has been on the polyamine group of proteins and ADAR1, a protein first isolated for its role in RNA editing.

Research into the polyamine group of proteins (putrescine, spermidine and spermine) has combined protein-binding studies with Z-DNA detection antibodies. These experiments have shown that these natural polyamines can bind to sequences of the type (dG-dC)n, (Thomas *et al.*, 1991), d(CG_3), (Ohishi *et al.*, 1991, 1996) and poly(dG-m_5dC).poly(dG-m_5dC), (Thomas *et al.*, 1988a & 1988b). The main feature of polyamine binding to any of the above sequences is that they seem to be capable of provoking the Z-conformation as opposed to stabilising an already Z-form DNA.

ADAR1 or DRADA1 (double-stranded RNA adenosine deaminase) is an enzyme first discovered for its role in pre-mRNA editing. It has the ability to deaminate adenosine to inosine, which is then read as guanosine. Editing of pre-mRNA thus allows variant proteins to be produced from the same DNA coding sequence. ADAR1 isolated from chicken lungs has now also been shown to bind specifically to Z-DNA (Herbert *et al.*, 1995a & 1995b). Within the protein two Z-DNA binding motifs have been isolated, Z-alpha and Z-beta (Herbert *et al.*, 1997 & Berger *et al.*, 1998), which interact to form a functional DNA binding site. Not only do these domains seem to recognise potential Z-forming sequences, but also seem to

be capable of then flipping them into the Z-conformation. This is even the case for the sequence $d(TA)_3$. A model has been proposed to explain these seemingly disparate roles of the same protein; Z-DNA binding motifs may target ADAR1 to regions of negative supercoiling (which stabilises Z-DNA). Supercoiling accompanies the trail of an actively transcribing RNA polymerase, and in this sense, RNA editing enzymes could thus be targeted to freshly transcribed regions, where they could take effect before other mechanisms such as RNA splicing and subsequent translation, can take effect (Herbert *et al.*, 1998).

Structural characterisation of the Z-alpha domain has revealed a winged helix-turn-helix motif, placing it in the class of proteins that contains histone H5 (Herbert & Rich, 1999; Schade *et al.*, 1999a & 1999b; Schwartz *et al.*, 1999a & 1999b). This domain has recently been characterised not only as a Z-DNA binding molecule, but also as a double-stranded Z-RNA binding molecule. Interestingly, Z-alpha has been shown by spectroscopic analysis, to induce a slow transition from the right handed A-conformation to the Z-conformation in the duplex $r(CG)_6$. A speculative role for the Z-alpha domain in targeting RNA viruses is discussed (Brown *et al.*, 2000). Problems exist however, in proposing a viable *in vivo* model(s) for the role of Z-DNA in light of various artefacts associated with the analysis of Z-DNA. For example, some antibody binding studies, (such as those that detected Z-DNA in the polytene chromosomes of *Drosophila*), have used acetic acid as a fixative (Hill & Stollar, 1983) which removes histones from DNA resulting in a reservoir of free energy that could be used for the promotion of Z-DNA regions (review; Hill, 1991). This was re-addressed in experiments that varied the acetic acid concentration and additionally used frozen sections prepared in ethanol with no acetic acid present. These results still indicated that Z-DNA formation correlates with the transcription puffs seen in polytene chromosomes but not in the resting state. However, by the author's own admission, these results did show some variance due to the conditions used (Lancillotti *et al.*, 1985 & 1987). Pardue *et al.*, (1983) used immunofluorescence to show effectively the same phenomenon– that Z-DNA correlates with transcriptionally active regions of DNA (Pardue, *et al.*, 1983). In a different type of assay, biotin-labeled anti-Z monoclonal antibodies were incubated with permeablised mammalian cell nuclei. The antibodies can diffuse within the nucleus over a small concentration

range. Radiolabelled streptavidin detects the bound antibody indicating that Z-DNA is associated with transcriptional activity, but to a lesser extent with replication. Topoisomerase I also dissolves the transcription-Z-DNA correlation (Wittig *et al.*, 1991). Using the same technique, Wittig *et al* looked at transcription of the human c-*myc* gene, using a laser to crosslink the antibodies to DNA. Subsequent restriction analysis revealed that the majority of Z-forming regions were centred on the promoter. Intriguingly, once the cells start to differentiate and *c-myc* transcription is down–regulated, the Z-DNA level goes to undetectable levels within 30-60 minutes (Wittig *et al.*, 1992).

Additionally, some proteins are known to cause DNA bending (eg: CAP) and thus the problem exists as to whether the structure being observed is a cause of the sequence context directly, or a consequence of non-specific protein-induced bending (El Hassan & Calladine, 1998). Another example is that of HMG1 and 2 proteins, found to bind brominated synthetic Z-form oligos of the type poly(dG-dC).poly(dG-dC), yet not capable of binding Z-DNA in supercoiled plasmids (Christen *et al.*, 1990; Rohner *et al.*, 1990). It has not yet been demonstrated whether the C5 –methylated version of the above sequence shows the same affinity with the HMG1 and 2 proteins. The ability of histones H1 and H5 to bind alternative sequences has also been questioned. With a set of tester oligomers it was found that H1/H5 bound Z-form DNA more readily than B-form, with H5 binding much stronger to the Z-sequence than H1 (Mura & Stollar, 1984). While observations such as these may reflect the ability of the lysine rich histones to enable various conformational changes, and thus have an effect on chromatin condensation in regions of sequence bias, the authors have not pursued this aspect.

Other evidence relating to chromatin structure relates to the ability of Z-DNA to be incorporated into nucleosomes (Miller *et al.*, 1985). This may be important in terms of gene regulation. Supercoil stabilised Z-form DNA consisting of a $d(CG/GC)_{12}$ plasmid insert was found to be completely refractory to nucleosome assembly. This is in stark contrast to the B-form of the same sequence, which readily allows nucleosome assembly (Casasnovas & Azorin, 1987). Similarly, reconstitution of nucleosomes onto a relaxed and supercoiled (G-C) rich template, shows that supercoiled (Z) DNA is not incorporated well. Micrococcal nuclease digests of

radiolabeled templates revealed that Z-form G-C sequences will sit outside the nucleosome core particles, whereas B-form G-C sequences are readily incorporated at the centre (Garner & Felsenfeld, 1987).

And finally, some phospholipid binding proteins that also bind Z-DNA were found only to do so because of a structural similarity to a phospholipid binding region (review: Herbert & Rich, 1996). Others, have used chemicals such as $[Co(NH_3)_6^{3+}]$ to induce Z-DNA at unrealistic concentrations and concluded a physiological relevance (Ma *et al.*, 1995). It seems the only consistent correlation however, is that a transcribing polymerase causing an increase in superhelical density behind itself, which may help stabilise Z-DNA, however transiently (Liu & Wang, 1987). In this respect, the only real function of Z-DNA could be to provide breathing space for a transcribing polymerase and to deflect nearby polymerases from approaching too soon. It is certainly true that Z-DNA is highly immunogenic, due to the high degree of exposure of the bases in this structure) as shown by many examples of a broad range of non-specific protein binding (review: van Holde & Zlatanova, 1994) and in effect such a barrage of temporary non-specific (or if proven otherwise, specific) protein binding could be simply to prevent re-transcription of such a region. Finally, in terms of me^5C, stabilisation of Z-DNA by methylation seems to override effects due to supercoiling (Casasnovas *et al.*, 1987, 1989), the function of which can be widely speculated on - in terms of transcription, replication and repair mechanisms; the reality being an immense lack of hard evidence.

1.4.2 A-DNA

It is more than reasonable to imagine a situation where the three conformations of DNA exist in a dynamic equilibrium with each other, yet can be stabilised in one form or another dependent on the situation. For example, we have seen that the B to Z- transition can be promoted and stabilised by DNA supercoiling, methylation, bromination and / or protein binding. Similar stabilising forces may also come into play for the transition to A-form DNA.

Sequences that are CG rich DNA are quite versatile and there does not need to be alternation of sequence for A-DNA to form. The dodecamer d(CCCCCGCGGGGG), has been crystallised as the first example of a complete turn

of A-DNA (Verdaguer *et al.*, 1991). Additional A-DNA studies have included analysis of sequences under unusual conditions. For example, Vorlickova *et al.*, (1991), demonstrated that the sequence poly(dI-dC) is capable of forming Z-DNA in concentrated (NaCl + $NiCl_2$) solution, yet can form A-DNA in trifluoroethanol solution. Additionally, some early studies, using transient electric dichroism, demonstrated that by varying alcohol/water concentrations for a poly[d(G-C)] fragment, it could be induced to form either A- or Z- DNA. The degree of hydration of both major and minor grooves was found to differ between both B- and A-forms, A-DNA being more heavily hydrated; the suggestion being that this may have some bearing on whatever role A-DNA may have *in vivo* (Crothers *et al.*, 1990). Single crystal structure analysis of oligonucleotide sequences provides more information regarding local variations in helix structure for a given sequence than can X-ray diffraction. In this respect, a tetramer d(iodo-CCGG) has been crystallised as a double stranded A-DNA helix (Conner *et al.*, 1982).The problem with some of these structural analyses is that the formation of A- or Z-DNA is not strictly confined to the base stacking preferences of a particular sequence. It also depends in some cases, on extrinsic factors/experimental conditions, which have altered the behaviour of the molecules. For example the use of extremely high salt and trifluoroethanol in some studies have played a large role in the production of not only A-DNA analysis but Z-DNA analysis also.

Methylation has also been found to influence the transition to A-form as it has for the B->Z transition. Mooers *et al.,* (1995) have solved the single-crystal X-ray structures of two different hexanucleotides: the alternating d(Gm5CGm5CGC) sequence and the non-alternating d(Gm5CCGGC) sequence, which both form A-DNA. Methylation of these sequences was found to confer stability to A-DNA, which in other cases has been provided by high salt or alcohol conditions. Therefore, CpG methylation may provide the physiological lever between the different conformational isomers in some sequence contexts.

Tippin *et al.*, (1997a) has found that the sequence d(CCGGGCCCGG) could be crystallised as A-DNA. When methylated or brominated at different cytosine residues, it also forms A-DNA, but with different influences on DNA bending, the structures fall into different orthorhombic groups. In addition to methylation,

spermine binding also promotes the A-conformation. Given that methylation and polyamines are so central to influencing gene expression, it seems reasonable to suppose that A-DNA may have a role *in vivo* in such processes.

Interestingly, Tippin *et al* (1997b), in an accompanying paper, have reconfirmed the promotion / stabilisation that methylation confers on A-DNA. They demonstrate that the alternating decamer d(GCGCGCGCGC) is expected to form, and is crystallised as Z-DNA. When this sequence is methylated to form d(Gm^{5}CGm^{5}CGCGCGC), this structure no longer crystallises as Z- but as A-form DNA. This is the first example of a sequence that when methylated forms A-DNA from the Z-conformation. From this study it was determined that 10 bp is the critical length upon which the handedness of d(CG)n type sequences is determined in the crystalline state (Tippin *et al*, 1997b). This occurs only after double methylation of the sequence - a feature consistant with the results of Mooers *et al.*, (1995), who found that the sequence d(GCGCGC) required methylation of at least one of the cytosines for it to be crystallised as A-DNA. This provides a situation in which the same sequence can be resolved into all B-, A-, and Z-forms. It had been observed that the hydration patterns of B- and A- DNA are very different. Mayer-Jung *et al.*, (1998) have focused on this issue by analysing the hydration patterns around methylated CpG steps in three high-resolution crystal structures of A-DNA decamers. In certain structural and sequence contexts, the methylated cytosine base is more hydrated than the unmodified form. The water molecules are coordinated to the methyl group through the formation of C-H$^{\cdots}$O bonds. The A-DNA structure also has magnesium cations bound to its major groove in a manner distinct from that of the unmethylated form. The suggestion is that methylated cytosine residues could be recognised by protein or DNA polar residues via their tightly bound water molecules (Mayer-Jung *et al.*, 1998).

Further to the above, a recent review article has speculated similarly; that bent / curved DNA that mimics a negative supercoil (Z-DNA) or a right-handed superhelical writhe (A-DNA) can help organise local chromatin infrastructure by directing some of the first contacts between cis-acting elements and their cognate binding factors (Ohyama, 2001).

1.5 MeCP binding proteins and DNA Recognition: A role for DNA Structure?

The significance of unusual forms of DNA to the methylation process is unknown. However, information is rapidly gathering about the protein complexes that attach to methylated DNA sites. The process by which methylated DNA is rendered transcriptionally inert is via the binding of methylation specific binding proteins such as MeCP1 and MeCP2 (review: Meehan *et al.*, 1993). MeCP1 can recognise and bind strongly to fifteen or more symmetrically methylated CpG dinucleotides (Meehan *et al.*, 1989) whereas MeCP2 recognises a single methyl CpG pair (Lewis *et al.*, 1992). Recent evidence provided by two independent groups, has shown that MeCP2 interacts strongly with a corepression complex containing mSin3A, a transcriptional repressor, and histone deacetylases HDAC1 and HDAC2 (Nan *et al.*, 1998, Jones *et al.*, 1998). Thus, it is via the convergence of two global mechanisms of gene regulation, methylation and histone deacetylation, that a repressive chromatin structure mediated by MeCP2, is established. MeCP2 has been found to contain an 85 amino acid long methyl-binding domain (MBD), which is responsible for binding of one or more symmetrically methylated CpGs. It also contains two short motifs that are also found in proteins that bind to the minor groove of AT-rich DNA. This is a non-specific DNA binding activity as the methyl group of cytosine is known to lie in the major groove of DNA, and competition experiments reveal that this activity is not required for methyl-specific binding (Nan *et al.*, 1993). It is not clear from these studies whether MeCPs can recognise and bind to A-form DNA. If methylation promotes the formation of A-DNA *in vivo* as it does *in vitro*, then it may be some feature specific to the nature of A-DNA that methyl-binding proteins such as MeCP2 can recognise and bind to. In support of this hypothesis, the affinity of MeCP1 for methylated DNA was challenged with different competitor DNA and RNA sequences: dsRNA (poly I.polyC), dsDNA poly (dI.dC) and poly d(I-C).d(I-C). Neither single (not shown) nor double stranded RNA competed with methylated DNA for MeCP1 complex binding, whereas both of the other two competitors could successfully compete for MeCP1 binding (figure 1.7; R. Meehan, unpublished). The poly (dI.dC) sequence has been shown by circular dichroism spectroscopy to form A-

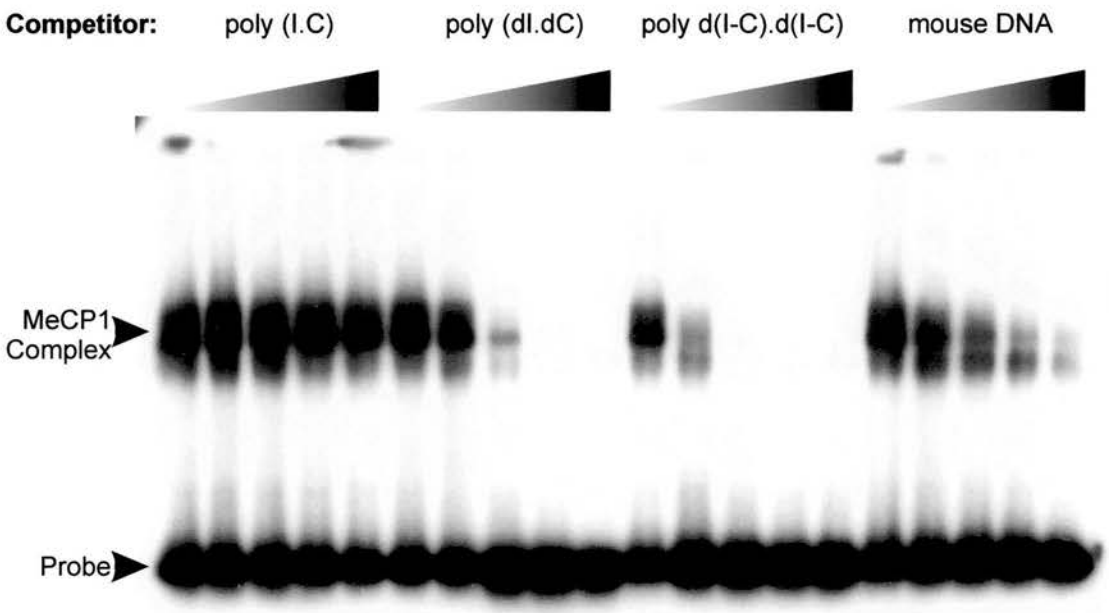


Figure 1.7 A-form DNA successfully competes with methylated DNA for MeCP1 binding. Competition experiment for rodent liver MeCP1 activity. MeCP1 activity can be effectively competed by poly (dI.dC) or poly d(I-C).d(I-C) but not by poly (I.C) competitors. The range of competitor used was: 0.2-2 μg. Mouse DNA competitor is shown for the purposes of comparison. (Unpublished data supplied by Dr. R. Meehan).

DNA (Vorlíčková & Sági, 1991). This has been confirmed by Raman spectroscopy (Fabriciova *et al.,* 1998). Further to this, it has also been shown that poly-(dI-dC) can be stabilised in the A-form on binding to berberine chloride at more physiological conditions than the studies mentioned above (Kumar *et al.,* 1992). It is possible that MeCP1 binding also has the same effect.

This experiment (figure 1.7) demonstrates that at least MeCP1 may require as a substrate, A-DNA or some feature of it. For example, recent evidence supports water molecule coordination to methylated cytosines in A-DNA (Mayer-Jung *et al.*, 1998), a feature that may aid in their recognition by MeCPs. Other methyl binding proteins could possibly also recognise A-form structures. This may be a universal feature of CpG methylated DNA. On the otherhand, this could provide a potential mechanism whereby other MeCPs are deflected from MeCP1 binding sites. It is possible that differences in CpG methylation densities can promote alternative DNA structures. It may be these methylation-induced changes in structure that help determine where and when different methyl-binding proteins bind to different methylation sites. The possible formation of A-DNA at methylated CpGs is investigated in this thesis.

1.6 Methyl-CpG Binding Proteins: the family.

Over the last ten years, a growing family of methyl-binding proteins has emerged from initially biochemical and subsequently database analysis. MeCP2 (Methyl-CpG binding protein 2) is the family prototype (Meehan *et al.*, 1989). Subsequently followed MBD1, MBD2, MBD3 and MBD4 (Hendrich *et al.*, 2001). These are all vertebrate methyl-binding proteins. They vary significantly in 3' end structure, but they commonly possess a methyl-CpG binding domain (MBD), see figure 1.8 *A* and 1.8 *B*.

Based on the hypothesis that methylated DNA may exert its effects via chromatin-associated proteins, the large multisubunit MeCP1 complex was identified. MeCP1 was isolated from band-shift assays using nuclear extracts as having specific binding to CpG methylated DNA. At least 12 symmetrically methylated CpG pairs are required for MeCP1 complex binding (Meehan *et al.*, 1989). The chromatin associated nuclear protein, MeCP2, was next identified. In contrast to the MeCP1

A

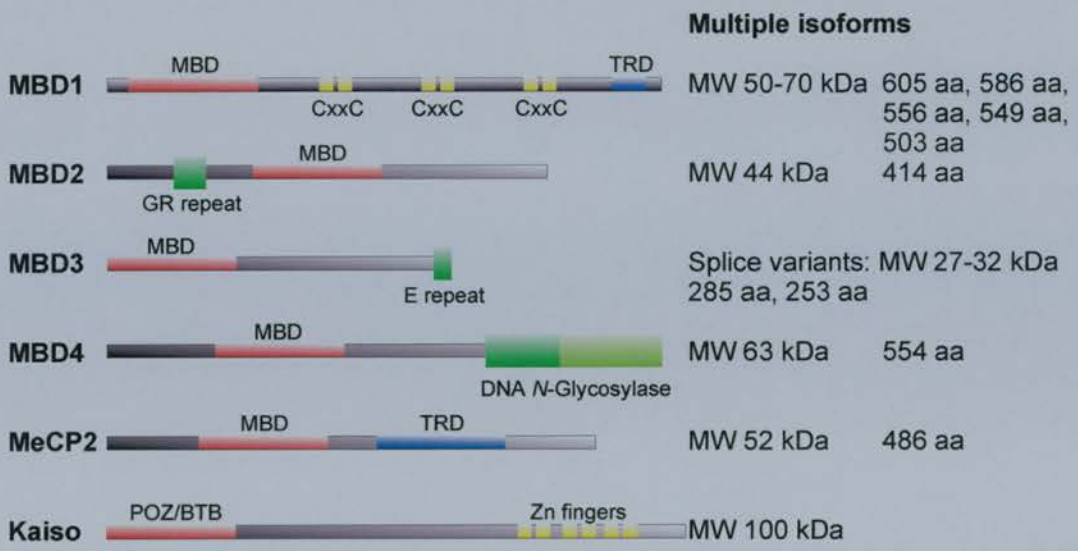
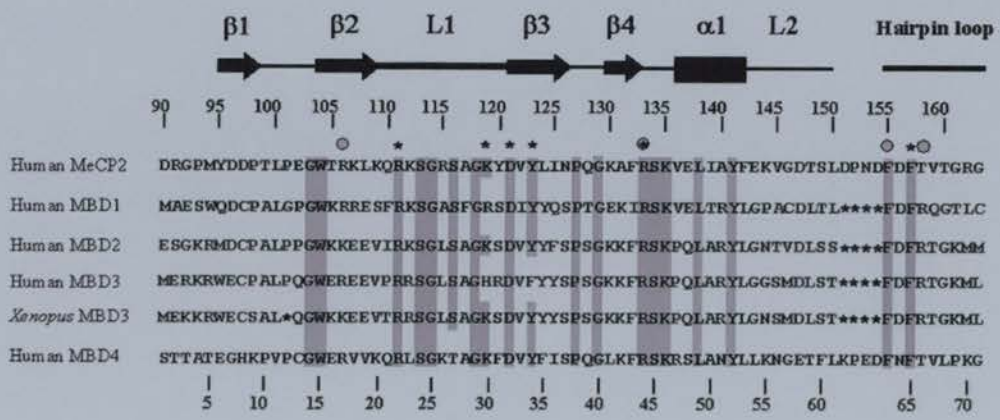


Figure 1.8 *A*, The methyl binding proteins. This figure depicts the organisation of the mammalian MBD containing protein family members in addition to Kaiso (not drawn to scale). The various members of the MBD family involved in transcriptional repression are depicted (adapted from Wade & Wolffe, 2001). Kaiso is a DNA methylation-dependant transcriptional repressor with links to the cell surface via p120 catenin (Prokhortchouk *et al.*, 2001). MBD, methyl-binding domain. TRD, transcriptional repression domain. CxxC, repeated motifs that can be alternatively spliced in the cDNA. E repeats, glycine repeats. Zn, zinc fingers indicate DNA binding domains. DNA *N*-Glycosylase, a domain responsible for demethylation activity. POZ/BTB, a Krüppel-like DNA binding domain. ***B*, Sequence alignment of the MBD from the methyl binding proteins.** Positions of β-strands (arrows), loops (thick lines), and the α-helix (rectangle) defined by the solution structures of MeCP2 and MBD1, are indicated above the alignment. MeCP2 numbering is located above the sequence. General numbering is located below the MBDs. Conserved residues are shaded, with those essential for meCpG binding indicated by an asterisk. Four Rett syndrome mutants occuring in the MBD are indicated by grey circles.

B



complex, MeCP2 was found to be a single protein of ≈ 55 kDa, with the ability to monomerically bind to a single symmetrically methylated CpG site (Lewis *et al.*, 1992; Meehan *et al.,* 1992; Nan *et al.*, 1993). The amino-terminal region contains a methyl-CpG binding domain (MBD), identified by deletion analysis. The C-terminal contains a transcriptional repression domain (TRD) required for transcriptional repression both *in vitro* and *in vivo*, in addition to several short motifs that bind to the minor groove of AT-rich sequences (Nan *et al.*, 1997, Jones *et al.*, 1998). These sequences are thought to contribute to a general non-specific binding of MeCP2 to DNA (Nan *et al.*, 1993). MeCP2 causes repression in a chromatin context. Both the MeCP2 protein and the isolated MBD bind to nucleosomes associated with methylated DNA (Nan *et al.,* 1993; Chandler *et al.*, 1999). Repression is also time-dependent (Kass *et al.*, 1997) requiring chromatin assembly before complete repression can be achieved. Subsequent analysis found that the TRD domain of MeCP2 can stably associate with the transcriptional co-repressor Sin3A and histone deacetylases HDAC1 and HDAC2. Moreover, in both mammalian cells and *Xenopus* oocytes, histone deacetylase inhibitors can partially relieve the repression exerted by the MeCP2–TRD at methylated gene promoters (Nan *et al.*, 1998; Jones *et al.*, 1998). More recently, it has been found that MeCP2 also interacts with two co-repressors, c-Ski and N-CoR, in addition to mSin3A. Antibodies directed against c-Ski and Sno (ski-related novel gene) shows colocalisation with MeCP2 and also abolishes Gal4-MeCP2 mediated transcriptional repression (Kokura *et al.,* 2001). The regulatory implications of these interactions are discussed later.

The solution structure of the MBD of rat MeCP2 (Wakefield *et al.*, 1999) and the solution structure of MBD1 (Ohki *et al.,* 2001) have now been determined. Figure 1.9 illustrates the structure of the MBD domain from both proteins. The fold of the MBD of both proteins is very similar. The strands *β2* and *β3* and the flexible loop show a high degree of similarity. The MBD is essentially a wedge shaped structure, one face of which has a series of hydrophobic basic residues that interact with DNA. For MeCP2, three of these solvent exposed residues (Tyrosine 123, Isoleucine 125 and Alanine 131) are likely to interact with the methyl groups in the major groove of DNA (Wakefield et al., 1999). Mutation of any of these residues reduces the binding efficiency of MeCP2 (Free *et al.*, 2001). In MBD1, only one of these positions

i

ii

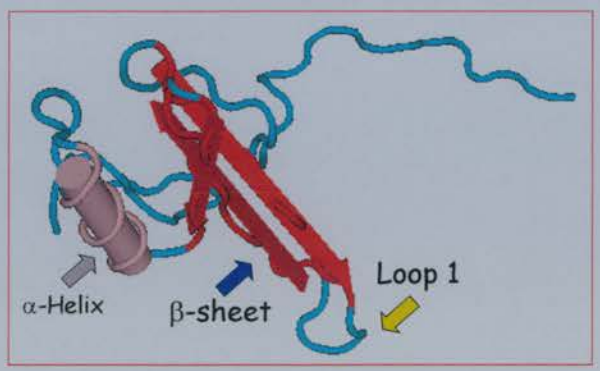


Figure 1.9 *A* The structure of the MBD domain from MeCP2. The MBD protein sequence is shown in ***i*** above. The MBD, as shown in ***ii*** above is wedge shaped, with one face comprising a β–sheet, the other consisting of an α-helix and hairpin loop. The vertex of the wedge is extended by a long loop between two of the β–strands and contains several basic residues vital for binding DNA. Binding DNA changes the resonance of several residues in this loop, in β–strands 2-4 and in the α–helix. These residues define the surface of a wedge that is likely to interact with DNA. Structural flexibility of the loop is of vital importance in this interaction. Three solvent exposed hydrophobic residues are proposed to interact with the methyl groups in the major groove. (Adapted from Wakefield *et al.*, 1999). Image contributed by Dr. R. Meehan.

i

ii

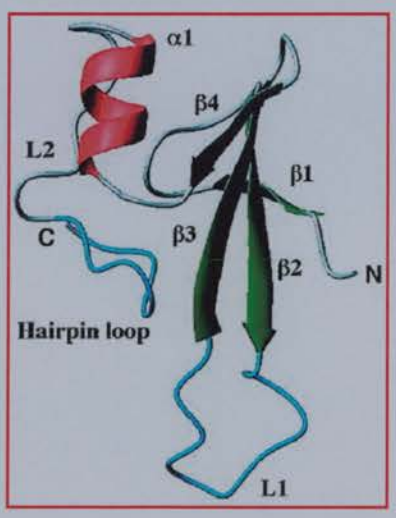


Figure 1.9 *B* The solution structure of the MBD of MBD1. The MBD is illustrated in complex with DNA in ***i*** above and the different domains illustrated in ***ii,*** above right. The MBD folds into an structure almost identical to that of the MBD of MeCP2. That is, an α/β sandwich structure comprising a layer of twisted β-sheet, backed by another layer formed by the α1-helix and a hairpin loop at the C terminus. Loop L1 undergoes major structural rearrangements upon binding DNA. Its flexible loop conformation adopts a well defined hairpin-like structure. The N– and C- termini are on opposite faces of the molecule. (Adapted from Ohki *et al.*, 2001).

contains a hydrophobic amino acid (tyrosine 34) and is absolutely conserved in all members of the MBD family that bind methylated DNA. Mutation of this residue completely abolishes binding by MBD1. It is likely that the hydrophobic residues on strand *β3* contact the methyl groups in the major groove of DNA and the charged residues in strand *β2*, *β4* and the flexible loop interact with the DNA backbone (review: Wade, 2001).

Underlining the importance of such MBD mutations *in vivo* has recently been uncovered. Patients who suffer from the neurodevelopmental disorder Rett Syndrome have mutations in the MeCP2 gene (Amir *et al.*, 1999). Since then, more than 90 different mutations clustering in the MBD and TRD, are present in more than 400 Rett Syndrome patients (Amir *et al.*, Buyse *et al.*, Bienvenu *et al.*, Cheadle *et al.*, Hampson *et al.*, Huppke *et al.*, Kim & Cook, Obata *et al.*, Xiang *et al.*, 2000). This is illustrated in figure 1.10. MeCP2 expression is highest in the postnatal brain consistent with a role in the mammalian central nervous system. Confirmation of the role in Rett's syndrome comes from MeCP2 null mice and brain specific deletion mutant mice. Specifically, mice deleted of the MeCP2 gene by the Cre-loxP recombination system at embryonic day (E) 12 have resulted in a phenotype identical to that of the null mutation. The results of this study also indicate that the role of MeCP2 is not limited to the immature brain, but becomes crucial in mature neurons (Chen *et al.*, 2001; Guy *et al.*, 2001).

The MBD sequence from MeCP2 was used in an EST database search to try and identify other proteins with the same structural motif, which might also bind to methylated DNA. This led to the identification of PCM1 (Cross *et al.*, 1997), later renamed MBD1. MBD1 can bind to and suppress transcription from a methylated template. MBD1 is a 70 kDa protein containing an N-terminal MBD. It is *not* a component of the MeCP1 complex as had been originally thought (Ng *et al.*, 2000). It also contains 2-3 cysteine-rich motifs (CxxC motifs) that are also present in DNA methyltransferase protein 1 (DNMT1) and the mammalian trithorax-like protein HRX. Alternative splicing determines the exact number of the CxxC motifs present. Interestingly, a second splicing event exists in mouse brain in the third exon, resulting in loss of the C-terminal half of the MBD, and another third splicing event that replaces the terminal 20 amino acids with a completely different 44 amino-acid

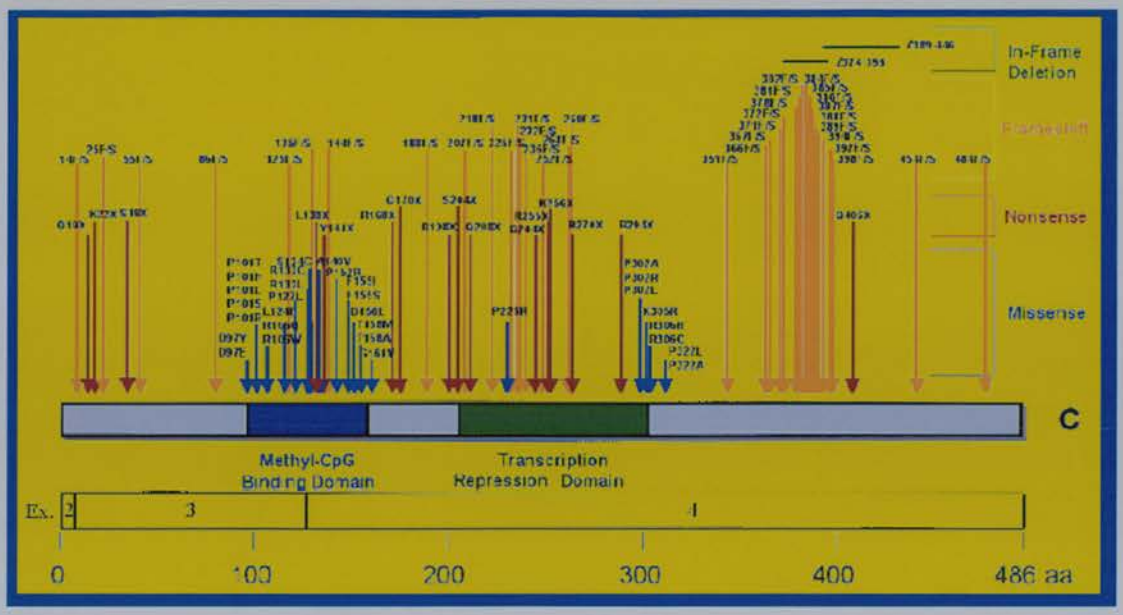


Figure 1.10 Mutations in MeCP2 found in Rett Syndrome patients. Mutations in MeCP2 corresponding to the neurodevelopmental disorder Rett Syndrome have been identified in 76% of the sporadic cases and in 45% of the familial cases of the disease. The amount and location of mutations can contribute to the severity of the disease (Van den Veyver *et al.*, 2000). MeCP2 is abundantly expressed in the brain in addition to other somatic tissues, giving support to the idea that mutations in MeCP2 may affect human development in a neurological phenotype (Tate *et al.*, 1996). The missense mutations (R106W, R133C, F155S, T158M) within the methyl-binding domain impairs selectivity for methylated DNA. The nonsense mutations (L138X, R168X, E235X, R255X, R270X, V288X, R294X) truncate all or some of the transcriptional repression domain (TRD). These affect the ability to repress transcription and have decreased levels of stability *in vivo*. Two missense mutations, one in the TRD (R306C) and one in the C-terminus (E397K), have no noticible effects on MeCP2 function (Yusufzai & Wolffe, 2000). Image adapted from Amir *et al.*, 1999).

terminus. Altogether there are at least five alternatively spliced variants of human MBD1 mRNA, with less in mouse (Hendrich *et al.*, 1999). The relevance of these alternative splice variants in biological terms is not yet understood, but not thought important for DNA binding at least (Cross *et al.*, 1997).

Like MeCP2, MBD1 localises to euchromatin with additional high concentrations at pericentric chromatin on mitotic chromosomes (Fujita *et al.*, 1999; Ng *et al.*, 2000). MBD1 associates remain unknown. Antisera specific for MTA1, HDAC1 and SAP30 do not deplete MBD1 suggesting that it is not part of the chromatin reorganising Mi-2/NuRD and Sin3 complxes, but is sensitive to deacetylase inhibitors. The effects of MBD1 and its splice variants on methylated DNA differ in their affects on transcription in transient assays (Fujita *et al.*, 1999). The solution structure of the MBD domain of MBD1 has also been solved (Ohki *et al.*, 1999) showing remarkable similarity to MeCP2. Interestingly, of the three hydrophobic residues predicted to bind methyl groups in MeCP2, only one of these is conserved (Wakefield *et al.*, 1999). Although no evidence exists, it is possible that these different structural organisations of the MBD allow interactions with different types of DNA structure perpetrated by different combinations of meCpGs.

Several more genes were subsequently identified from EST database searches. These were MBD2, MBD3 and MBD4. Figure 1.8 *B* shows the amino acid sequence alignment of the different MBD containing complexes. Among the methyl-binding proteins, it appears that the MBD of MeCP2 and MBD4 share most similarity, while MBD1, MBD2 and MBD3 constitute a separate subfamily (Hendrich & Bird, 1998). All contain a conserved intron, linking them evolutionarily. All of the *mbd* genes are expressed in all murine tissues analysed (Hendrich & Bird, 1998) and are not expressed to any appreciable level in ES cells, as expected, as they do not require / possess methylation. MBD1, 2 and 4 (but not MBD3), all bind specifically to methylated DNA *in vitro*. Consistent with this, MBD-GFP constructs, apart from MBD3, also localise to heavily methylated satellite DNA regions (Hendrich & Bird, 1998).

MBD2 can exist as MBD2a, a 43.5kDa protein or MBD2b, a 29.1 kDa protein that lacks the amino terminal 140 amino acids. This is dependent on where initiation of translation takes place. Alternative splicing can also result in a variety of nonsense

transcripts. MBD2b can repress transcription from a methylated template and in this respect behaves similarly to MeCP2. It is also physically associated with HDAC1 and Sin3A in mammalian cells, making it a good candidate for MeCP1 complex involvement (Boeke *et al.*, 2000). It also interacts with RbA p48/p46 by coimmunoprecipitation analysis (Ng & Bird, 1999). However, in the same analysis, it was not found in this case to interact with mSin3A and Mi-2/NURD complexes (Ng *et al.*, 2000). The exact components of the MeCP1 complex have since been delineated in a separate study. Feng and Zhang (2001) have purified the 1MDa MeCP1 complex from HeLa nuclear extracts, to homogeneity. It contains 10 major polypeptides. They include 90% of the total [MBD2], two uncharacterised polypeptides of 66 and 68kD, and components of the NuRD complex; (HDAC1/2, RbAp46/48, the histone deacetylase core; Mi2, a SWI2/SNF2 type helicase/ATPase involved in the chromatin remodelling function of the NuRD complex; and MTA2, MBD3). This complex of proteins has been found to bind to, remodel and deacetylate methylated nucleosomes *in vitro* (Feng & Zhang, 2001). Recently, a candidate for one of the unknown MeCP1 components has been found in a yeast two-hybrid screen using MBD2 as bait. A novel zinc finger protein, MIZF of between 60 and 68 kDa was found to interact with MBD2 (Sekimata *et al.,* 2001). *In vitro* binding assays and IPP experiments confirmed this. MIZF contains 7 zinc fingers in its' C-terminal, 4 of which are required for MBD2 binding. Interestingly, MIZF mRNA is expressed in all human tissues and cell lines examined, but the MIZF protein localises diffusely in the nucleoplasmic region, while MBD2 localises to major satellite regions. The two proteins do co-localise in some regions of the nucleus, however. MIZF classically behaves as a transcriptional repressor that interacts with HDAC1, the repression being attenuated by trichostatin A and is completely dependent on interacting with MBD2. With specific relevance to this project, MIZF interacts with some DNA sequences but not all. The subset of sequences that it recognises may be a factor in silencing specific genes with the aid of MBD2 and NuRD complex components. Initial evidence also suggests an interaction with MBD3, another member of the MeCP1 / NuRD complex (Sekimata *et al.,* 2001).

Controversially, MBD2b has been suggested to have demethylase activity (Bhattacharya *et al.*,1999), adding weight to the idea that demethylation is not solely a passive mechanism: ie: that methylation levels are reduced by replication in the

absence of a DNA methyltransferase. However much uncertainty remains over the validity of these results, as attempts to reproduce them by other groups have not been successful (Ng *et al.*, 1999; Wade *et al.*, 1999).

MBD3 also has splicing variants. The predominant form is a 32 kDa protein with a high degree of homology (70% identity) to MBD2b. Despite the high degree of homology it is clear that MBD2 and MBD3 in mice, play distinctive but interacting roles. In mice, MBD3 fails to bind methylated DNA yet MBD3(-/-) mice die in early embryogenesis. In contrast, MBD2 binds methylated DNA, yet MBD2(-/-) mice survive and are fertile but show defective maternal nurturing. MBD2 and MBD3 do interact genetically, the evidence favouring a model whereby MBD2 may recruit the MBD3 containing Mi-2/NuRD complex to DNA (Hendrich *et al.,* 2001). The less frequent form has a major deletion in the MBD domain (MBDΔ). Both isoforms exist in human and mouse (Hendrich *et al*, 1999; Wade *et al.*, 1999; Zhang *et al.*, 1999). *Xenopus laevis* has an additional 'long form' variant containing a 20 amino acid insertion in the MBD. MBD3Δ and MBD long form do not bind methylated DNA (Wade *et al.,* 1999). Mammalian MBD3 will only bind methylated DNA *in vitro* very weakly. However, both the *Xenopus* (Wade *et al.*, 1999) and Zebrafish (Macleod *et al.*, 1999) MBD3 homologues, can bind methylated DNA. Despite a less acidic C-terminal, the xMBD3 has two conserved residues that are mutated in the mammalian variants. These may be responsible for this: a K/R30H mutation and another change to a Y at position 34 (Ohki *et al.*, 1999). These residues correspond to the same sites in MeCP2 that are mutated in some Rett's patients (Ballestar *et al.*, 2000; Yusufzai *et al.*, 2000). Additionally, these residues when mutated in MBD1, also abolish DNA binding (Ohki *et al.,* 1999). The function of MBD3 in these variant isoforms is not understood. As one possibility, it has been tested for demethylase activity but to date, none has been detected (Wade *et al.*, 1999). It is, as mentioned part of the Mi2 complex from HeLa cells (Feng & Zhang, 2001). This is also the case in *Xenopus* eggs (Wade *et al.,* 1998). However, both *Xenopus* and Zebrafish developmental programmes do not contain the global demethylation events characteristic of mammalian development, and therein may be clues to the difference in behaviour of the MBD components. In both these organisms a period of transcriptional silencing precedes their developmental programme and may be at this stage that the MBD

proteins are most active in gene repression (Stancheva & Meehan, 2000; Martin *et al*, 1999). An interesting paradox exists in that an MBD2/3 homolog, dMBD-like, exists in *Drosophila melanogaster,* which has little or no DNA methylation. The dMBD-like protein fails to bind methylated DNA, but possesses the biochemical properties consistent with roles as a component of a histone deacetylase-dependent corepressor complex similar to the vertebrate Mi-2 complex (Ballestar, *et al.*, 2001).This suggests a broader role in terms of chromatin repression mechanisms than simply binding to methylated DNA. It may also suggest that there is a specific pattern of DNA methylation that dMBD-like may recognise that has not been tested yet.

MBD4 is 62kDa and like the other MBDs, has splice variants, although no splicing event occurs within the MBD. MBD4-GFP fusions localise to densely methylated DNA sequences in mouse cells. MBD4 is also the only member of the MBD family not associated with histone deacetylase activity. Interestingly although the MBD of MBD4 has the highest homology to MeCP2, the MBD4 C-terminus has homology to bacterial DNA repair enzymes (Hendrich & Bird, 1998). These include the 8-oxoGA mispair-specific adenine glycosylase MutY of *E. coli,* and the photodimer-specific UV-endonuclease of *Micrococcus luteus*. Based on this similarity it was tested whether MBD4 has any repair activities directed to methyl-CpG sites. It was found that MBD4 can efficiently remove mispaired U or T residues from CpG sites *in vitro* and thus functions to decrease the mutation rates at highly mutagenic CpG sites (Hendrich *et al.*, 1999). It is thus a DNA repair enzyme specific for methylated DNA. This is strengthened by the isolation of mutations in the MBD4 gene from carcinomas with microsatellite instability, where DNA mismatch repair is defective (Bader *et al.,* 1999; Riccio *et al.,* 1999).

Apart from the above mentioned proteins which were identified due to (1) the ability to bind methylated DNA, (2) the presence of an MBD or (3) the ability to interact with another MeCP or component of the NuRD / MeCP1 complex, several other proteins have come to light. One of these is a BTB/POZ domain zinc finger transcription factor called *Kaiso* (Prokhortchouk *et al.,* 2001). It has structural similarity to a family of known *Drosophila* transcriptional regulators. It has now also been isolated from *Xenopus* embryos (Dr. Rusov, pers. comm.). *Kaiso* was first identified for its interaction with the catenin $p120^{ctn}$ protein. $p120^{ctn}$ is an Armadillo

repeat domain protein that interacts directly with E-cadherin, a transmembrane cell adhesion molecule. E-cadherin is involved in diverse functions, but $p120^{ctn}$ does not seem to be involved in these and may be a downstream effector of these pathways (Daniel *et al.,* 1999). *Kaiso* was then found to be capable of binding specifically to DNA with a consensus sequence containing multiple meCpG sites. It shows no sequence similarity to the known MBD proteins, yet associates *in vivo* with HDAC1, and HDAC2 (Prokhortchouk *et al.,* 2001). This distinct preference for a consensus sequence adds weight to an hypothesis of a sequence based coding system for an array of MeCPs, recognising methylated CpGs by their density and sequence context. These aspects are discussed further in the Discussion, Chapter 6.

It has been the search for candidate genes for tumourigenesis that had led to the identification of a cluster of proteins that may also be involved in methyl-mediated repression. These proteins were not identified in any of the previous MBD screens. In B-Cell Chronic Lymphocytic Leukemia a large region of chromosome 13q14 is deleted. Analysis of this region led to the identification of three candidate genes for leukomogenesis. These are *CLLD6, CLLD7* and *CLLD8*. The CLLD8 protein contains an MBD domain, a preSET domain and a SET domain, suggesting an involvement with methylation mediated transcriptional repression. The SET domain was also first identified in the *Drosophila* chromosomal regulator proteins; Su(var)3-9, Enhancer of zeste (E[z]), and Trithorax (TRX) (Tschiersch *et al.,* 1994) but is present in a huge range of transcriptional regulators from a variety of species (Hobert *et al.,* 1996). CLLD8 is most similar to the Su(var)3-9 subgroup. The preSET domain, also known as the SAC domain (SET domain-associated cysteine-rich domain), is probably involved in chromosome binding (Huang *et al.,* 1998). SUV39H1, the human homologue of *Drosophila* Su(var)3-9 has also a SET and preSET domain. It has recently been reported to be a histone methyltransferase (Rea *et al.,* 2000). CLLD8, containing an MBD in addition to SET/preSET domains, is also likely to be associated with methyl-mediated transcriptional repression, which also involves histone methylation (Mabuchi *et al.,* 2001). Several more histone methyltransferases have now been discovered. Along with the histone acetyltransferases, the methyltrasferases play a key role in regulating gene expression in a chromatin complex (review: Zhang & Reinberg, 2001; Jenuwein & Allis, 2001).

The large unstructured tails of the histone proteins can also be phosphorylated, ubiquitinated and brominated at specific amino acid residues leading to the hypothesis that there is a 'histone code' that operates analogously to the DNA code. The combinatorial nature of the different modifications seems to provide functional and heritable epigenetic information on a particular gene sequence or region of DNA. The complexity is increased if the array of modifications on a particular histone tail can combine its coded information with adjacent histone tail modifications (Jenuwein & Allis, 2001). Specific examples include (a) the parent-specific complementary patterns of histone H3-K9 and H3-K4 methylation at the Prader-Willi syndrome imprinting centre (Xin *et al.,* 2001) and (b) the involvement of HMT activity in Rb-mediated transcriptional repression, which also utilises HDACs for this purpose (Vandel *et al.,* 2001). This last example reveals what may be a complex interplay of the histone acetylation and methylation machinery responding to the effects of methylation mediated repression.

It is now impossible to say whether DNA sequence, structural topology or histones alone are ultimately responsible for, or can directly control gene expression. What must exist is a circular interweaving pathway between these components maintaining the heritable nature of these modifications in a form of cellular memory. Figure 1.11 is a summary of some of the components of the methylation machinery that may interact at a given point in time to enable this system of regulation to be perpetuated. What follows is an investigation into the role that CpG methylation of the primary DNA sequence may have in the realm of epigenetic phenomena.

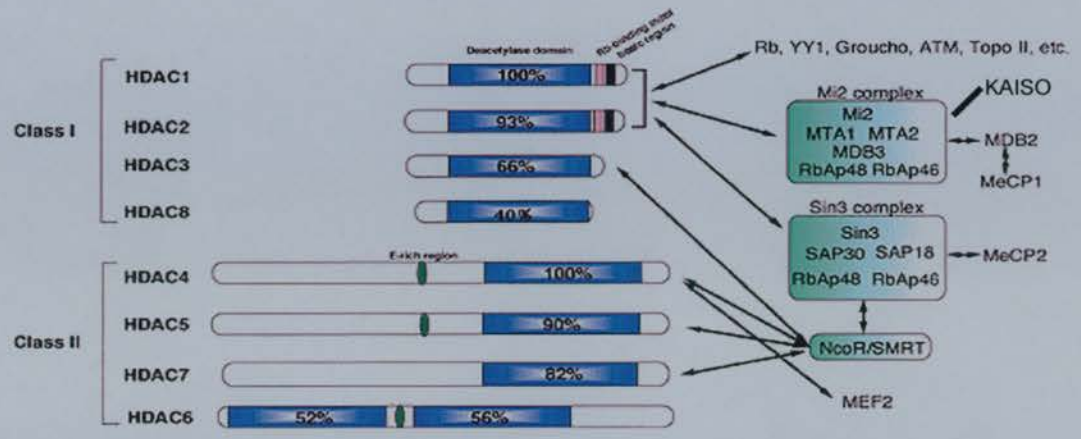


Figure 1.11 The methylation machinery interacts with chromatin remodelling complexes containing a variety of Histone Deacetylases. This diagram illustrates how the methyl-binding proteins interact with chromatin remodelling complexes, which in turn interact with a variety of histone deacetylases (HDACs). The potential for differrent interactions of these componants illustrates the maleability of regulation at genes under this type of regulation. Targeting of these complexes is mediated by the methyl-DNA binding proteins. (Image received from Dr. R. Meehan).

1.7 Aims of Thesis

The Aims of this thesis are centred on an investigation into the effect that CpG methylation of DNA may have on DNA structure. The initial premise was to either devise a novel *in vitro* assay or to modify an existing one, with the purpose of investigating the effects of CpG methylation on DNA structure. The object was then to pursue the relevance of any potential methylation-induced structural changes in a functional context.

Here I demonstrate, that CpG methylation in a variety of DNA substrates can cause significant structural alterations. These alterations occur in a specific sequence context and are consistent with the formation of A-form DNA. An assay for methylation induced DNA structural changes was developed using Benzonase, a nuclease with a distinct preference for the digestion of A-DNA. For this, a number of 'tester' sequences, and a variety of specific gene sequences were methylated and subjected to Benzonase digestion.

The results reveal the possibility that a DNA structural code is induced by differential methylation. These patterns may be individually recognised by an array of methyl-binding proteins, thus providing complex patterns of gene regulation at specific sites in DNA.

Chapter 2: Materials & Methods

Unless otherwise stated, all molar concentrations in this section are final concentrations and chemicals were obtained from BDH

2.1 Materials

2.1.1 List of Suppliers

ABgene: ABgene House, Blenheim Road, Epsom, Surrey. KT19 9AP England. U.K. Tel: +44(0)1372 723456.

Ambion: Ambion (Europe) Ltd, Spitfire Close, Ermine Business Park, Huntingdon, Cambridgeshire PE29 6XY.England. UK.Tel 0800 138 1836

Amersham-Pharmacia Biotech UK Ltd. Amersham-Pharmacia Biotech UK Ltd., Amersham Place , Little Chalfont, Buckinghamshire HP7 9NA. England. UK. Customer Services: 0800 515 313.

Applied Biosystems: Applied Biosystems, 7 Kingsland Grange, Woolston Warrington, Cheshire WA1 7SR. United Kingdom. Phone: 44 1925 825650

BDH: Merck House, Poole, Dorset. BH15 1TD. England. UK. Tel: 01202 669700.

Beckman: Oakley Court, Kingsmead Business Park, London Road, High Wycombe, Buckinghamshire. HP11 1JU. England. UK. Tel: 44-1494-441181

BIO-RAD: Bio-Rad House, Maylands Avenue, Hemel Hempstead Hertfordshire HP2 7TD. England. UK. Toll free: 1-(800) 4 BIORAD.

Difco Laboratories Ltd.: PO Box 14B, Central Avenue, West Molesey, Surrey, KT8 2SE. London. UK. Phone: 0181 979 9951.

Fisher Scientific UK Ltd: Bishop Meadow Road, Loughborough Leicster, LE11 5RG. England. UK. Phone: 01509 231166.

Greiner Labortechnik: Unit 5, Stroudwater Business Park Brunel Way, Stonehouse, Gloucestershire GL103SX. England. UK. Phone: 01453 825 255.

Helena-Biosciences Ltd.: Colima Avenue, Sunderland Enterprise Park, Sunderland, SR5-3XB. Phone: 0191-549 6064.

Hybaid Ltd: Action Court Ashford Road, Ashford Middlesex. TW15 1XB. England. UK. Phone: 01784-425000.

Invitrogen Ltd. Life Technologies Ltd: Unit 3 Fountain Drive, Inchinnan Business Park, Paisley. PA4-9RF. Scotland. UK. Phone: 0141-814 6100.

MBI-Fermentas Ltd. Helena Biosciences Ltd., Colima Avenue, Sunderland Enterprise Park, Sunderland, SR5-3XB. Phone: 0191-549 6064.

Molecular Probes Europe BV. PoortGebouw, Rijnsburgerweg 10, 2333 AA Leiden. The Netherlands. Phone: +31-71-5233378.

MWG Biotech AG (UK) Ltd.: Mill Court, Featherstone Road, Wolverhampton Mill South Milton Keynes. MK12 5RD. England. UK. Phone: 01908 525 500.

New England Biolabs (UK) Ltd: 73 Knowl Piece, Wilbury Way, Hitchin, Hertfordshire. SG4 OTY. England. U.K. Freephone: 0800-318486.

Novagen: Boulevard Industrial Park, Padge Road, Beeston, Nottingham. NG9 2JR. UK. Freephone: 0800-622935.

Pierce Ltd.: Perbio Science UK Ltd., Century House, High Street, Tattenhall, Cheshire. CH3 9RJ. Tel: 01829-771 744.

Princeton Separations: CP Instrument Co. Ltd., Unit 3, The Shires, Shirehill Industrial Estates, Saffron Walden, UK. CB11 3AN. Phone: 01719-581-321.

Promega UK. Delta House, Chilworth Research Centre, Southampton. SO16 7NS. England. UK. Phone: 023-8076 0225.

Qbiogene. Oakland Park Road, Oakbank Business Park. Livingston. EH53 OTG. England. UK. Phone: 0800-328-8401.

Qiagen Ltd: Boundary Court, Gatwick Road, Crawley, West Sussex. RH10-9AX. UK. Orders: 01293-422-922.

Roche Diagnostics Ltd: Bell Lane, Lewes, East Sussex. BN7 1LG. England. UK. Phone: 01273-480444.

Sigma-Aldrich Company Ltd: Poole, Dorset. England. UK. Phone: 01202-733144.

Sorval / Heraes Brands: Kendro Laboratory Products Ltd. Stortford Hall Park Bishop's Stortford. Hertfordshire. CM23 5GZ. UK. Phone: 01279-827700.

Stratagene Europe Corporate Office: Gebouw California, Hogehilweg 15, 1101 CB Amsterdam Zuidoost, The Netherlands. Phone Orders (UK) 0800-585-370.

Thermo-Hybaid Ltd. Medical Supply Co. Ltd. Damastown, Mulhuddart. Dublin 15. Republic of Ireland. Phone: 00-353-31-82224 222

UVP Inc. (Europe Office): Ultra Violet Products Ltd., Unit 1, Nuffield Road, Trinity Hall Farm Estate, Cambridge. CB4 ITG. England, UK. Phone: 01223-420022.

2.1.2 Stock Solutions List

Acrylamide, 30% (v/v) 29:1 (acryl:bis-acryl): This solution was used for all non-denaturing polyacrylamide gels. It was made by mixing three parts of 40% acrylamide (Fischer) with one part ddH_2O.

Acrylamide, 30% (v/v) 19:1 (acryl:bis-acryl): This solution was used for all denaturing polyacrylamide-urea gels. It was made by mixing three parts of 40% acrylamide (Fischer) with one part ddH_2O.

Ampicillin (Sigma) solution: This antibiotic solution was made by dissolving ampicillin in ddH_2O to a final concentration of 100 mg/ml. The solution was filter sterilised, aliquoted to 1ml in eppendorf tubes and stored at $-20^{\circ}C$. The working concentration is typically between 50 and 100 μg/ml of bacterial growth media.

APS (10%): Ammonium persulphate was used for catalysing polymerization of polyacrylamide gels. It was prepared by dissolving 1 g of APS in 10 ml of ddH_2O and stored at $4^{\circ}C$ for up to two weeks.

APS (25%): Ammonium persulphate was used at 25% (w/v) for catalysing the polymerisation of denaturing PAGE / urea gels. It was prepared by dissolving 2.5 g of APS in 10 ml of ddH_2O and stored at $4^{\circ}C$ for up to two weeks.

Bandshift Buffer: For Protein-DNA interactions by PAGE. A 5X solution contains 100 mM Hepes, 5 mM EDTA pH 8.0, 15 mM $MgCl_2$, 50 mM β-mercaptoethanol, 50 % glycerol and 0.5 % Triton X-100.

BCA-200 protein assay kit (PIERCE): This kit was used for the determination of protein concentrations using spectrographic standards of bovine serum albumin concentrations, according to the manufacturer's instructions.

BPC (Buffered phenol-chloroform): BPC was used for DNA extraction/purification methods. To prepare, 25 parts of buffered (TE-saturated) phenol, 24 parts of chloroform and 1 part of isoamylalcohol was mixed by vigorous vortexing, left until phases separated and stored in the dark at 4°C.

BPB (bromophenol blue): This tracking dye for gel electrophoresis was prepared by dissolving 1 g of BPB in 10 ml ddH_2O. The solution was aliquoted to one ml eppendorfs and stored at –20°C.

Buffer A (Midipreps): 25mM Tris-Cl, pH 8, 10 mM EDTA, 50 mM glucose.

Buffer B (pET6hMBD purification): A solution containing 50 mM NaCl, 50 mM sodium phosphate, pH 7, 10% glycerol and 0.1% Triton X-100 was made by dissolving all components in a flask in 500 ml ddH_2O. The solution was then made up to 1L with ddH_2O.

Buffer Z: For HIV-1 integrase purification protocol. A stock solution contains 1 M NaCl, 20 mM Hepes pH 7.5, 2 mM β-mercaptoethanol, 0.3 mg/ml lysozyme and 5 % glycerol. The solution is made to 500 ml with ddH_2O.

$CaCl_2$ (1M): A 1L solution of calcium chloride was prepared by dissolving 219g of $CaCl_2$-$6H_20$ in 800ml ddH_2O. Once dissolved the volume was adjusted to 1 litre with ddH_2O and sterilised by autoclaving. The solution was stored at room temperature.

Coomassie Blue staining solution: For the staining of SDS-PAGE gels. To prepare a 1L solution, 2.5g of 'Coomassie Brilliant Blue R-250' was dissolved in a mixture of 450 ml methanol, 100 ml acetic acid and 400 ml ddH_2O. The volume was adjusted to 1L with ddH_2O and the solution stored at room temperature.

Destain I solution: For the destaining of Coomassie stained SDS-PAGE gels. A stock solution is made by mixing 500 ml methanol, 400 ml of acetic acid and 100 ml ddH_2O.

Destain II solution: For the destaining and re-sizing of gels destained with 'Destain II' solution detailed above. A stock solution is made by mixing 200 ml of methanol, 100 ml of acetic acid and 700 ml of ddH_2O.

DNA polymerase I large (Klenow) fragment: For use in all PCR amplification protocols, was obtained from New England Biolabs, and used according to manufacturer's instructions.

DNase I Solution: Deoxyribonuclease I (ABgene) solution was used for the enzymatic cleavage of DNA and prepared as follows: 10 mg of DNase I was resuspended with 0.1 M Tris, pH 7.5 and aliquoted into eppendorf tubes to a final concentration of 10 mg/ml The working concentration used was typically at 1-100 µg/ml with 5 mM $MgCl_2$ pH 7.0 - 8.5.

DTT (1M): A 20ml solution of 1,4-dithio-DL-threitol (DTT, MW=154.25) was made by dissolving 3.085g of DTT in 20 ml of 10mM sodium acetate (pH 5.2). The solution was divided into 1 ml aliquots and filter sterilised through a 0.2 µm filter. Aliquots were stored at -20°C.

Gel Loading Buffer (6 ×): This was used for sample loading in polyacrylamide gel electrophoresis: 60 µl 30% (w/v) Ficoll 400, 30 µl 10 × TBE, 3 µl 2% (w/v) XC, 3 µl 2% (w/v) BPB and 4 µl ddH_2O was mixed together.

EDTA (Na^2+EDTA) 0.5M, pH 8.0. A one litre solution is made by dissolving 186.12 g of disodium ethylenediaminetetracetate $-2H_2O$ (Mw=372.24) in 800 ml H_2O and stirred vigorously on a magnetic stirrer. The pH was adjusted to pH 8.0 with NaOH (approx. 20g NaOH pellets), made up to 1 L with H_2O and sterilised by autoclaving.

EtBr (sigma)10 mg/ml: A stock solution of ethidium bromide, for the staining of DNA in agarose and polyacrylamide gels, was prepared by dissolving 1g of EtBr in

100 ml ddH_2O. The solution was stored in a dark container at 4°C. For gel staining: 0.5 – 1 ug per ml of agarose solution was typically used.

Gel Extraction buffer: Buffer for the extraction of DNA from polyacrylamide gels. A stock solution was made of 0.3 M NaCl, 20 mM Tris-Cl pH 7.51 and 1 mM EDTA to a total volume of 50 ml.

Gel loading buffer for agarose and polyacrylamide gels was prepared by mixing 500 µl 10 × TBE, 400 µl 30% (w/v) Ficoll 400, 80 µl 2% (w/v) BPB and 20 µl 2% (w/v) XC.

GTE buffer: For use in alkaline lysis preparation of plasmid DNA: 50 mM glucose, 25 mM Tris-HCl pH 7.4 and 10 mM EDTA was dissolved in ddH_2O and made to a final volume of 50ml.

Hepes (1M): A stock solution is made by dissolving 23.83 g in 80 ml of ddH_2O. The pH is adjusted to 7.5 with 0.5 N NaOH.

IPTG (100mM): A 50ml solution of IPTG was made by dissolving isopropyl b-D-thiogalactopyranoside (IPTG, Mw=283.3) in 40ml H_2O. The volume was adjusted to 50ml with H_2O, divided into 5ml aliquots, filter sterilised and stored at –20°C.

KOAc (0.1 M) Buffer: Acidic potassium acetate solution was made as follows:

(a) A solution of 0.2M acetic acid was made by mixing 11.55ml of glacial acetic acid in 500ml H_2O and the volume adjusted to 1L with ddH_2O.

(b) A solution of 0.2M potassium acetate was made by dissolving 19.62g of KOAc, (Mw=98.14) in 800ml ddH_2O and the volume adjusted to 1L with H_2O.

A quantity of each solution is mixed proportionately to give a specfic pH and made to a final volume of 100ml with ddH_2O. The quantities used are tabulated in Sambrook, 1997.

K Phosphate (0.1M) Buffer: A 200ml solution of 0.1M potassium phosphate was prepared as follows:

(a) A solution of 0.2M potassium phosphate, mono-sodium salt was prepared by dissolving 27.2g of KH_2PO_4 (Mw=136.09) in 500 ml H_2O and the volume adjusted to 1L wilth ddH_2O.

(b) A solution of 0.2M potassium phosphate, di-sodium salt was prepared by dissolving 34.8g of K_2HPO_4 (Mw=174.18) in 500 ml ddH_2O and the volume adjusted to 1L with ddH_2O.

A quantity of each solution is mixed proportionately to give a specific pH and made to a final volume of 200 ml with ddH_2O. The quantities used are tabulated in Sambrook, 1997.

LB (Luria broth) medium: A solution was made by dissolving 10 g bactotryptone, 5 g bacto-yeast extract and 10 g NaCl in ddH_2O to 1L and subsequently autoclaved.

LB Agar: 15 g/l of agar was added to LB medium (see above) and autoclaved.
If required, the appropriate antibiotic was added immediately before pouring the plates, when the liquefied Agar had cooled to at least 50°C.

Lysis buffer (plasmid preparation): 200 mM NaOH and 1% (w/v) SDS was dissolved in ddH_2O to 50ml final volume.

Lysozyme: Stock solution is made by dissolving 10 mg/ml in 10mM Tris-Cl, pH 8.0.

$MgCl_2$ (1M Magnesium Chloride): A 1L solution of $MgCl_2$ was made by dissolving 203.31g $MgCl_2$ –$6H_2O$, Mw=238.31) in 800ml H_2O. The volume was adjusted to 1L with ddH_2O and sterilised by autoclaving.

NaCl (5M): A 1L solution of 5M Sodium Chloride was made by dissolving 292.2g (NaCl, Mw=58.44) in 800ml H_2O. The volume was adjusted to 1L with ddH_2O, sterilised by autoclaving and stored at room temperature.

NaCl (1M): A 500ml solution was made by mixing 100ml of 5M NaCl (see above) with 400ml of ddH_2O.

NaOAc (3M): A 1L solution of sodium acetate was made by dissolving 408.24g NaOAc-$3H_2O$ (Mw=136.08) in 800ml of ddH_2O. The pH was adjusted to 5.2 with glacial acetic acid and made up to 1L with ddH_2O. The solution was sterilised by autoclaving and stored at room temperature.

NaOH (1N): This is equivalent to 1M NaOH solution and is made by dissolving 40 g in 1000 ml ddH_2O.

Na Phosphate Buffer (0.1M): A 200ml solution of 0.1M sodium phosphate was prepared as follows:

(a) 0.2M sodium phosphate, mono-sodium salt was prepared by dissolving 27.6g of NaH_2PO_4-$1H_2O$ (Mw=138) in 500ml of ddH_2O and the solution adjusted to 1L with H_2O.

(b) 0.2M sodium phosphate, di-sodium salt was prepared by dissolving 53.62g of Na_2HPO_4 –$7H_2O$ (Mw=268.1) in 500ml of ddH_2O and the solution adjusted to 1L with H_2O.

A 200ml solution of the desired pH is prepared by mixing a quantity of solutions (a) and (b) together to a final volume of 200ml with ddH_2O The exact amounts are according to Sambrook, 1997.

dNTPs: Deoxyribonucleotide triphosphates: dATP, dCTP, dTTP and dGTP; (Pharmacia Biotech) were diluted to 10 mM each, in ddH_2O and stored at -20°C.

PEG solution A solution of polyethylene glycol was prepared for use in alkaline lysis preparation of plasmid DNA: 20% (w/v) PEG-6000 and 2.5M NaCl were dissolved in ddH_2O to 50 ml.

Phenol (Buffered pH 8.0): 250 g of solid phenol (Fluka) was dissolved overnight at room temperature in 127 ml of 2 M Tris-Cl (pH 7.5). Following mixing, the phenol

was centrifuged at 2000 x *g* for 1 minute, and the aqueous bottom layer was discarded. Then, 55 ml of 2 M Tris (pH 8.0), 13.75 ml of m-cresol, 550 µl of β-mercaptoethanol and 275 mg of 8-hydroxyquinone was added. The solution was mixed and divided into aliquots, which were centrifuged to separate the buffer and phenol layers prior to storage at 4°C in the dark.

PBS (10X PBS Buffer): A solution was made by dissolving 80g $NaCl_2$, 2g KCl, 26.8g $Na_2HPO_4 -7H_2O$ and 2.4g KH_2PO_4 in 800ml H_2O. The pH was adjusted to 7.4 with HCl and the volume increased to 1L with H2O. The solution was sterilised by autoclaving and stored at room temperature.

PBS/2 % dried milk: 1 g of Marvel is dissolved in 50 ml of 1X PBS.

PMSF: A phenylmethyl sulphonyl fluoride (PMSF) solution was made by dissolving solid PMSF in isopropanol to a final concentration of 250mM. The solution was maintained soluble at > 30°C.

Ponceau S solution: An 100 ml solution was made by dissolving 2g of Ponceau S (sigma), 30g of trichloroacetic acid and 30 g of sulfosalicylic acid in 80 ml of ddH_2O. The volume was adjusted to 100 ml with ddH_2O, and the solution stored at room temperature.

Proteinase K was obtained from NBL at 50 mg/ml and stored at -20 °C.

Radioactive nucleotides: α-^{32}P-dCTP and γ-^{32}P-dATP (Amersham-Pharmacia) were purchased at 0.37 MBq/µl = 5000 Ci/mmol = 10 mCi/ml.

RNase A: Bovine pancreatic ribonuclease A (BCL) was dissolved at 10 mg/ml in 0.01 M sodium acetate pH 5.2 and boiled for 15 minutes to destroy any contaminating DNase I activity. After the RNase had cooled slowly to room temperature 0.1 volumes of 1 M Tris-Cl (pH 7.4) were added.

SDS {10% (w/v)} An 100 ml solution was prepared by dissolving 10 g in 90 ml ddH_2O. To aid the process the solution was heated to 68°C. The pH was adjusted to 7.2 with HCl (approx. 5 µl), the volume adjusted to 100 ml with ddH_2O and the solution stored at room temperature.

SDS-PAGE (10X) running buffer: This was made by dissolving 10g of SDS, 30.3 g of Tris and 144.1 g glycine in 800 ml ddH_2O. The volume was adjusted to 1 L with ddH_2O and the solution stored at room temperature.

SDS-PAGE (2X) sample buffer: Aliquots of an 100 ml stock solution were made by mixing 10 ml of 1.5 M Tris (pH 6.8), 6 ml of 20% SDS, 30 ml of glycerol, 15 ml of β-mercaptoethanol and 1.8g of bromophenol blue. The volume was adjusted to 100 ml with ddH_2O, aliquoted into 10 ml stock solutions and stored at –20°C. Working solutions were stored at 4°C.

SOB medium: A 1 L solution is made by mixing 20g of Bacto-Tryptone (Dipco), 5g of Bacto-Yeast extract (Dipco), 0.5g of NaCl and 2.5 ml of 1M KCl in 900 ml of H_2O. The pH was adjusted to 7.0 with 10M NaOH (approx. 100 µl) and ddH_2O added to 990 ml. The solution was sterilised by autoclaving and stored at room temperature. Before use, 5 ml of sterile 1M $MgCl_2$ and 5 ml of sterile 1M $MgSO_4$ was added to the medium.

SOC medium: The solution is identical to SOB medium (see above) except that it additionally contains 20 ml of sterile 1M glucose.

SSC (20X) Buffer, pH 7.0: This was prepared by dissolving 175.3g NaCl and 88.2g sodium citrate-$2H_2O$ in 800ml H_2O. The pH was adjusted to 7.0 with HCl and the solution made to a final volume of 1L with ddH_2O. The solution was autoclaved.

TAE (50 X) Buffer: A 1 L solution of Tris-acetate buffer was made by dissolving 242g Tris in 500ml H_2O. To this, 100 ml of 0.5M Na_2EDTA (pH 8.0) and 57.1 ml of

glacial acetic acid was added. The volume was adjusted to 1L with ddH_2O and stored at room temperature.

TB (terrific broth) growth medium was made with 12 g bactotryptone, 24 g bacto-yeast extract, 4 ml glycerol, 17 mM KH_2PO_4, 55 mM K_2HPO_4 and 20mM glucose, dissolved in ddH_2O to a final vol. of 1L and subsequently autoclaved.

TBE (10 X) Buffer: A 1 L solution Tris-borate buffer was made by dissolving 108 g of Tris base, 55 g orthoboric acid and 9.3 g EDTA in ddH_2O to a final volume of 1L The solution was autoclaved and stored at room temperature.

TBS: Used for Western Blots. A solution is made by mixing 25 ml of 1 M stock of Tris-Cl pH 8.0 (50 mM final) and 15 ml of 5M NaCl (150 mM final) to a final volume of 500 ml in ddH_2O.

TCA (100% Trichloroacetic acid): 500g of TCA was added to 227ml H_20 to make an 100% solution.

TE (1X): 10ml of 1M Tris (pH 8.0, 7.6 or 7.4) and 2ml of 0.5M Na_2EDTA (pH 8.0) was added to 800ml ddH_2O, mixed and the volume adjusted to 1L. The solution was autoclaved and stored at room temperature.

Transformation Buffer ('TB'): A 500 ml solution of 'TB' contains 1.5 g of PIPES (10 mM final), 1.1g of $CaCl_2$ (15 mM final) and 9.3g of KCl (250 mM final). The pH was adjusted to 6.7 with KOH and then $MnCl_2$ added to 55 mM (5.44g). The solution was sterile filtered at 4°C. and stored in 100 ml aliquots at –20°C.

Tris-HCl (1M tris(hydroxymethyl)aminomethane): A solution was made by dissolving 121.14 g of Tris base (Mw=121.14) in 800ml of ddH_2O and the pH adjusted to the desired value with concentrated HCl.

e.g: pH 7.4 : Approx 70ml, pH 7.6 : Approx 60ml, pH 8.0 : Approx 42ml.

The volume was adjusted to 1L with H_2O, sterilised by autoclaving and stored at room temperature.

XC (Xylene Cyanol) 2% (w/v) for tracking of buffer fronts in gel electrophoresis: 1 g was dissolved in ddH_2O to a final vol. of 50ml.

2.2. Methods:

The following methods are those used in Results, Chapters 3, 4, 5 and Chapter 7.

A list of Bacterial host strains and plasmid vectors used is provided in table 2.1.

Bacterial Host Strains	**Genotype**
XL1-blue	F'::Tn*10 proA*$^+$*B*$^+$ *lacI*^{q}*Δ(lacZ)M15/recA1 endA1 gyrA96* (Nal^r) *thi hsdR17* ($r_K^-m_K^+$) *glnV44 relA1 lac*
JM110	*rpsL thr leu thi hsd*R17 (r_K^-, m_K^+) *lacY galK galT ara tonA tsx dam dcm supE*44 *Δ(lac-proAB)* [F' *traD*36 *proAB lacI*^{q}ZΔM15
BL21(DE3)pLysS	F-, *ompT, hsdS*$_B$ (r_B-, m_B-), *dcm, gal,* λ(DE3), pLysS (Cm^r)
Plasmids	**Contents**
pBS SK(-/+) (chip30)	An MeCP2 ChIP assay pull down sequence (Appendix) Obtained from Dr. I. Stancheva
pBR322	Standard vector, no DNA insert (Chapter 3). Commercial.
pGEM9zF(-)	pGEM vector with 167 bp DNA insert (Chapter 3) Obtained from Dr. A. Pingoud.
pUC9 (α-globin)	pUC9 cloning vector with human α_2-globin DNA insert (Ch 3). Obtained from Prof. A. Bird
pBS KS(-)1 (β-globin)	The chicken β^A-globin gene (Chapter 4) Obtained from Dr. C. Davey
pET6hMBD	HIS tag protein expression vector with MeCP2-MBD (Ch5) Obtained from Dr. R. Meehan
pGEX-3T-MBD	GST tag protein expression vector with MeCP2-MBD. (Ch 5) Obtained from Dr. R. Meehan
pIN$^{1\text{-}288}$F185K/C280S	Soluble mutant of Human HIV-1 integrase pET expression vector (Ch 7). Obtained from Dr. R. Craigie

Table 2.1 List of commonly used Bacterial cells and plasmids in this thesis. XL1-blue cells were obtained from Stratagene, BL21 (DE3) pLysS cells from Novagen. JM110 cells were a gift from Dr. C. Davey.

2.2.1 Preparation of Competent Cells

This procedure was modified from a high efficiency transformation procedure detailed by Inoue *et al.*, 1990, and was performed on XL-1 Blue, JM110 and BL21(DE3)pLysS cells. Glycerol based cell stocks were stored at –80°C for long term storage. A cell scraping was made from a still frozen vial on ice and used to initiate a starter culture by inoculating 5 ml of LB medium. Antibiotics were added if required, and the cells grown overnight at 37°C. Control reactions were set up containing LB only to confirm that the LB was indeed an uncontaminated stock. The following day the 5ml culture was used to inoculate 250 ml of SOB medium. The culture was grown at 18°C to an $OD_{600} = 0.6$. The culture was then placed on ice for 10 minutes and then spun at 2500 x g (5000 rpm in a Sorvall GSA or 3000 in a Beckman J-6B centrifuge) for 10 minutes at 4°C. The cells were gently resuspended in 80 ml of ice cold transformation buffer and left on ice for 10 minutes. The cells were then spun again at 2500 x g for 10 minutes at 4°C. Cells were resuspended in 40 ml of ice cold transformation buffer. DMSO was added to a final concentration of 7% and the cells kept on ice for 10 minutes. Aliquots of 1 ml were made in eppendorf tubes placed in ice and the cells snap frozen with liquid nitrogen. Competent cells were stored at –80°C until required.

2.2.2 Transformation and growth of plasmid DNA in bacterial host strains

For transformation of plasmid DNA into a bacterial cells, 200 µl of freshly prepared competent cells were taken and allowed to thaw on ice. To these, 3 µl of DMSO was added and 15 ng to 200 ng of the desired plasmid DNA. These were mixed gently with a pipette and left on ice for 30 minutes. The cells were heat shocked at 42°C for 1.5 minutes in a heating block to promote uptake of plasmid DNA. To these 1ml of LB was added and the cells placed in a shaking incubator at 37°C for 1 hour. They were then spun down gently at 2000 rpm for 30 seconds and the excess LB removed with a pipette. The cells were resuspended in the remaining LB and aseptically plated onto antibiotic containing selective LB plates. Controls of non-transformed cells were also plated on antibiotic containing plates and LB only plates. Plates were incubated overnight at 37°C to allow colonies to grow.

2.2.3 Agarose Gel Electrophoresis

Agarose gels were prepared according to standard methods (Sambrook, 1997). Typically a 1% gel was made by dissolving 1g of agarose (sigma) in 100 ml of 1X TAE buffer, by heating the solution to boiling point. The solution is allowed to cool while being continuously stirred. EtBr is added to the desired concentration (0.5 μg/ml), the solution poured to a gel mould and allowed to set. For Low melting point gels, Seaplaque^{TM} agarose was used according to manufacturer's instructions.

2.2.4 Polyacrylamide Gel Electrophoresis

Polyacrylamide gels were prepared according to standard methods described in Sambrook, 1997. Typically a 12 % polyacrylamide gel for oligo purification was made by mixing 40 ml of 30 % acrylamide, 49.3 ml of ddH_20, 10 ml of 10X TBE and 0.7 ml of 10 % APS. The solution is mixed and 50 μl of Temed (Sigma) added. It is again mixed, poured carefully into the gel mould and allowed to set at room temperature. A 90 ml solution suitable for the preparation of two 5% PAGE gels for DNA or protein analysis was made by mixing 15 ml of 30 % acrylamide (29:1), 65 ml of ddH_20, 9 ml of 10X TBE, 1 ml of 10 % APS and 120 μl of Temed. Polymerisation is rapid and within 5 minutes. 5% PAGE gels for bandshift analysis were made to ½ the normal concentration of TBE (ie: 0.5X TBE) and run in the cold room in 0.5 X TBE Buffer.

2.2.5 Staining of Agarose and PAGE gels

After the gel run is complete, agarose and polyacrylamide gels were stained with either EtBr from a solution of 0.5 mg / ml. Typically, EtBr is added before agarose gel electrophoresis until the solution turns only very slightly pink when mixed. EtBr can also be added after agarose gel electrophoresis and always after PAGE gel running. This provides a more consistent staining pattern. The gel is placed in a dish with a lid and is surrounded by running buffer. EtBr is added to a final concentration of 0.5 μg/ml or until the solution is markedly red. After 20 minutes of agitation the solution is washed away. Replacement buffer without EtBr is added to destain the gel. After destaining, the gel is ready to be viewed under UV light which allows visualisation of sites in DNA where EtBr has intercalated.

Another method used is SYBR gold (Molecular Probes) staining. Where this is used is indicated in the main text. It is a fluorescent dye that can also be visualised under UV light but is more sensitive than EtBr. It detects dsDNA and ssDNA as well as RNA. SYBR-gold is diluted as 5 µl in 50 ml gel-staining buffer solution, which is the same as the running buffer. Application must be done in the dark as SYBR-gold is light sensitive.

2.2.6 Medium scale growth (midiprep) of plasmid DNA

A 10 ml culture was set up by picking a single colony from a transformation plate to 10 ml LB containing antibiotic (eg: ampicillin at 50 µg/ml). This starter culture is allowed to grow at 37°C overnight in a shaking incubator. The following day, 5 ml of the starter culture was transferred to a flask containing 200 ml of LB containing the appropriate antibiotic. The culture was allowed to grow at 37°C in a shaking incubator until it reached an OD_{600} of 0.8. The cells were then spun down in a Sorvall GSA centifuge at 4000 rpm for 10 minutes. The pellet was resuspended in 5 ml 'Buffer A' with lysozyme at 5 mg/ml and left on ice for 5 minutes. To this, 10 mls of 0.2 N NaOH and 0.1% SDS were added in that order. The solution was left on ice for no more than 5 minutes. 7.5 ml of 5M Potassium acetate pH 4.8, was then added and again the solution was left on ice for no more than 5 minutes. To the suspension 5 ml of chloroform was added and mixed well. The solution was centrifuged at 3,000 rpm for 5 minutes in a table top swing rotor. The supernatant was removed and an equal volume of isopropanol was added at room temperature. Immediately, the solution was centrifuged at 12,000 rpm for 10 minutes. The pellet was washed with 70% ethanol and again centrifuged at 12,000 rpm for 10 minutes. The pellet was dried at room temperature for 10 minutes. To this, 3 ml of TE was added with RnaseA (@ 20 µg/ml) and incubated at 37°C for 30 minutes. An equal volume of phenol was then added, the solution mixed and the aqueous and organic phases resolved by centrifugation at 3,000 rpm for 5 minutes. The aqueous layer was removed to an equal volume of phenol/chloroform and the mixture again centrifuged at 3,000 rpm for 5 minutes. The aqueous supernatant was added to an equal volume of chloroform, the solution mixed and centrifuged at 3,000 rpm for 5 minutes. The supernatant is taken and aliquoted into 2 ml eppendorf tubes with 800 µl in each

tube. To each tube, 1/10 volume of 3M Sodium acetate and 2 ½ to 3 volumes of ice-cold 100% Ethanol was added to precipitate the DNA. Each tube was placed at -20°C overnight to aid precipitation. The tubes were centrifuged in an eppendorf centrifuge at maximum speed, 13,000 rpm. The pellets were washed with 70% Ethanol and allowed to dry at room temperature for 10 minutes. The pellets were allowed to dissolve in 400 µl TE (for the total 200 ml culture), at 4°C overnight.
Samples were tested for purity and concentration by both spectrometry and agarose gel electrophoresis.

2.2.7 Small-scale growth (miniprep) of plasmid DNA.

A single colony is picked from an overnight transformation plate and used to inoculate 5-10 ml of Terrific Broth containing the appropriate antibiotic (eg: 50 µg/ml ampicillin). The starter culture was grown up overnight at 37°C. The cells were centrifuged at 13 K for 5 minutes and resuspended in 600 µl TELT. The resuspension was transferred to an eppendorf tube and 40 µl of 100 mg/ml of lysozyme added and left for 2 minutes. The suspension was boiled for 1.5 minutes, then cooled on ice for 10 minutes. Cell debris was pelleted by spinning the suspension for 15 minutes at 14 K in an eppendorf centrifuge. The supernatant was removed to a fresh tube and 0.6 of a volume (360 µl) of isopropanol was added. The solution was left on ice for 2 minutes and then spun out at 14 K for 20 - 30 minutes. The pellet was washed in 70% ethanol and allowed to dry in air for 15 minutes. The pellet was then resuspended in 50 µl of TE buffer and a sample removed for quantification by agarose gel electrophoresis.

2.2.8 Amplification of DNA by the Polymerase Chain Reaction (PCR).

Amplification of DNA by the polymerase chain reaction (PCR) is a technique used to produce large amounts of a specific DNA fragment. Forward and reverse primers are annealed to a target plasmid or dsDNA sequence and copies of the target sequence created by repeated rounds of primer annealing and extension of primers with DNA polymerase to create dsDNA products. Primer annealing and extension conditions determined according to standard methods described in Sambrook, 1997. Here, PCR is used to amplify regions of plasmid DNA useful for Benzonase, DNaseI

and/or Footprinting analysis. Table 2.2 lists the primers and primer sequences used for PCR analysis.

2.2.9 Dephosphorylation of 5'-ends of dsDNA for radioactive labelling

Dephosphorylation of 5' ends of dsDNA was required to prepare an efficient template for 5' labelling with γ-^{32}P-ATP. For marker DNA, approximately 2 µg of either pBR322/BsuRI (MBI) or pUC19/Hin61 were dephosphorylated with CIAP (MBI-fermentas) or Shrimp Alkaline Phosphatase (Amersham) in a 50 µl reaction volume according to methods described by each company. For Benzonase, DNaseI and Footprinting analysis, typically 2-10 µg of DNA is dephosphorylated.

2.2.10 Preparation of radiolabelled marker DNA

For denaturing PAGE-urea gels, a variety of markers were used. These were radiolabelled with ^{32}P-γ–ATP (Amersham). Commercial markers used were pBR322 cut with *Bsu*RI {*Hae*III} (MBI-fermentas) and pUC19 cut with *Msp*I {*Hpa*II} (MBI-fermentas). The pBR322 marker ranges from 8 bp to 587 bp and the pUC19 marker ranges from 26 bp to 501 bp. Therefore, they provide complementary information about different sized bands, for the same size range. Both have a concentration of 0.5 mg/ml. For each marker, a 10 µl sample was taken, dephosphorylated as described and subsequently radiolabelled. Labelling of between 10 and 100 pmoles of dsDNA was performed with ^{32}P-γ-ATP (Amersham) in a final volume of 50 µl, with 12 units of T4 polynucleotide kinase in 1X T4 PNK reaction buffer. Labelled markers were purified using CentriSpin-10 columns (Princeton-Separations) and stored for use at –20°C in a Perspex container. For gel analysis, a 2 µl sample was taken, diluted to 20 µl with Sequencing gel buffer and 5 µl of this mix loaded per lane as required.

Other markers used for PAGE-urea gels were generated from the DNA fragment being observed. For example to highlight specific regions of interest a sample of between 20 ng to 500 ng of 5'-labelled unmethylated DNA was kept aside prior to either Benzonase or DNaseI analysis. This fragment was digested with *Hin*61 {*Hha*I} (MBI-fermentas) to highlight 5'-GCGC-3' regions, or with *Hpa*II to highlight 5'-CCGG-3' regions in the dsDNA fragment.

Forward Primer	Reverse Primer	Plasmid	Product Use
BARTF ACTCATACTCTTCCTTTTTC	BARTR TAATGCGGTAGTTTATCACA	pBR322	(Chapter 3)
pGSp6F GCAGATTTAGGTGACACTATAGAATA	pGT7R TAATACGACTCACTATAGGGAGC	pGEM9zf(-)	(Chapter 3)
GloF2 TAGCATGCTCTAGACTGCAGGAAGC GAGGCTGGAGA	GloRev2 AAGGCCTTGCATGCGTCCCTAG CGAGGGAGAAGTC	pUC9(α_2-globin)	(Chapter 3)
Bglo(F) ATAAGCTTGATCTGGTGT	Bglo(R) AGCTCTGCAGCTCTATAC	pBS-KS(-)1(β^A-globin)	(Chapter 4)
Chip30(F) CCTAATGAGTGAGCTAA	Chip30(R) ACACTATAGAATACTCA	pBS-SK(-)1(chip30)	Appendix

Table 2.2 Primers for PCR generation of DNA fragments for structural analysis. Here primers used for the generation of DNA fragments by PCR are listed. All gene fragments produced are of the same seuence as those listed throughout the results section. That is, final products are the same whether generated by PCR or from Plasmid digests. The pBR322 product is first used in 3.1.2.2. The pGEM9zf(-) is first used in 3.1.2.4. pBS, refers to the pBLUESCRIPT plasmid and contains the chicken β^A-globin gene. The pUC9 plasmid contains the human α_2-globin gene. Chip 30 refers to a ChIP pull-down sequence obtained by fishing with MeCP2 in ES cells.

2.2.11 Radioactive labelling of substrate DNA for Benzonase, DNaseI and Footprinting analyses.

Radioactive labelling of DNA at the 5'-ends with ^{32}P-γ-ATP for Benzonase, DNaseI and footprinting analyses was essentially carried out as described for the preparation of radiolabelled markers. In some cases the reaction was scaled up to obtain a high incorporation rate of radiolabel. After labelling of the 5' ends of DNA, the DNA was purified by a CentriSpin-20 (Princeton) column and one of the 5' labelled ends of the fragment removed by restriction enzyme digestion. Centrispin –30 or –40 columns were used to remove the labelled end. In most cases removal was unnecessary as the cleaved DNA fragment was small enough to not interfere with structural analysis. If the restriction enzyme buffer was suitable, Benzonase digestion could be carried out directly. In most cases the DNA was either re-purified on a CentriSpin column or phenol-chloroform extracted, ethanol precipitated and resuspended in ddH_2O.

2.2.12 Preparation and Running of denaturing PAGE/urea Sequencing gels

Denaturing PAGE/urea gels were prepared in line with standard methods described in Sambrook, 1997 with some modifications made according to the apparatus used and improvements to the gel mix to improve image resolution. Four different concentrations of PAGE/urea were prepared as shown in Table 2.3 below.

Percentage	6 %	8 %	10 %	12 %
urea	187.43 g	187.43 g	187.43 g	187.43 g
40% Acrylamide (19:1)	69.94 ml	93.53 ml	117.26 ml	140.52 ml
10 X TBE	44.64 ml	44.64 ml	44.64 ml	44. 64 ml
ddH_2O	197.92 ml	174.33 m	150.59 ml	127.41 ml

Table 2.3 Denaturing PAGE/urea gel mix recipies. The above components were each mixed together and gently heated on a heating block to promote solubilisation of urea. The mass of each solution is approximately 500 g. For a single gel, 100 ml of solution is used is taken. To this 90 µl of 25 % APS and 125 µl of Temed is added. The solution is mixed thoroughly and used quickly as polymerisation is rapid. 20-30 ml of the solution is used to wash out gel pouring apparatus (syringes and tubing) and the rest is used to pour the gel.

Sequencing gel apparatus used was obtained by BIO-RAD and is the Sequi-GenII model. Gel pouring is through the base of the gel. A 14 well comb was added to the top of the gel, which is then sealed with clingfilm and allowed to set. For use, the gel

is preheated at 50 W for ½ hour to 1 hour until the temperature reaches a stable 50^0C. The wells are rinsed to remove urea prior to sample loading. The gel is again run at 50 W for 2 to 3 or more hours depending on the % acrylamide and the size of the DNA fragment.

2.2.13 Drying and Exposure of PAGE/urea gels

Denaturing PAGE-urea gels were typically run for between 2 and 3 hours. Once electrophoresis was complete, the gels were allowed to cool back to room temperature. The plates were separated carefully and the gel allowed to soak in a layer of 10 % acetic acid / 10 % methanol solution to leach out the urea. The acetic acid/methanol was added carefully to the surface of the gel, which was balanced over a drainage sink. After 15 minutes, the solution was tipped off and the gel surface rinsed, with several careful washes of ddH_2O. Two layers of 17 CHR paper (3MM) were cut to gel size and soaked briefly in ddH_2O. Each sheet was levered gently onto the gel surface. To ensure no air bubbles are present, the paper was flattened by rolling a long glass rod from one end of the gel to the other. Two larger 17 CHR sheets were used as backing and the gel flipped over and placed horizontally so that the glass plate holding the gel faced upwards. The plate was carefully lifted, while the 3M paper was removed from underneath along with the gel. The removed gel was covered with Saran ^{TM} wrap and dried on a gel drier for 30-40 minutes at 80^oC under air suction. The dried gel was then placed in a phosphorImager (Fugifilm) cassette with a white Fugifilm-screen and left overnight at room temperature. The following day, the screen was scanned in a phosphorImager (Fugifilm) according to manufacturer's instruction with 16 bit resolution, 500 dpi and a sensitivity of F1=1000. Scanned images were later reduced in size to TIFF and JPG files if this did not result in any loss of resolution.

2.2.14 SDS-PAGE for the denaturation and resolution of proteins.

Standard SDS-PAGE analysis was carried out for all protein expression and purification stages mentioned throughout the thesis.

Table 2.4 outlining the volumes of each component used to obtain different concentrations of separating gels for SDS-PAGE is provided below.

Percentage	7%	10%	12%	15%
ddH_2O	5.05 ml	4.05 ml	3.35 ml	2.35 ml
1.5 M Tris-Cl, pH 8.8	2.5 ml	2.5 ml	2.5 ml	2.5 ml
10% SDS	0.1 ml	0.1 ml	0.1 ml	0.1 ml
Acrylamide/Bisacrylamide (30%/0.8% w/v), 29:1	2.3 ml	3.3 ml	4.0 ml	5.0 ml
10% APS	0.05 ml	0.05 ml	0.05 ml	0.05 ml
TEMED	5 µl	5 µl	5 µl	5 µl
Total Monomer	10 ml	10 ml	10 ml	10 ml

Table 2.4 Separating gels for SDS-PAGE analysis. Separating gels are prepared to the desired concentration dependent on the range of protein separation required. Temed is added last to the reaction and mixed thoroughly. Gels are poured as described in Sambrook, 1997 and according to the manufacturers recommendations, in this case BIO-RAD. Separating gels are allowed to polymerise at room temperature and are overlayed with ddH_2O to aid the process. Once set the ddH_2) is poured off and a 4% stacking gel is overlayed and allowed to set. Stacking gels are made by mixing 3.3 ml of ddH_2O, 1.25 ml of 0.5 M Tris-Cl, pH 6.8, 50 µl of 10 % SDS, 670 µl of acrylamide/Bisacrylamide (30%/0.8% w/v), 25 µl of 10% APS and 15 µl of Temed.

2.2.15 Induction procedure for GST-tagged MBD protein

The following is an optimised protocol for the bacterial cell based expression of the rat GST tagged MBD protein. It was also applied to a series of purifications of the methylated DNA binding protein, *Kaiso.* These purifications were not used in this thesis but may be used by others. Many attempts were made to obtain pure and active protein based around standard methods of GST-fused protein induction. Four sets of 20 ml starter cultures containing LB with ampicillin at 50 µg/ml were grown up overnight at 37°C in a shaking incubator. These contained a scraping from a culture of BL21(DE3)pLysS cells transformed with pGEX-3T-MBD; a plasmid containing the MBD domain of the rat MeCP2 protein. The following day 15 mls of each culture were each transferred to a flask containing 1L of LB with ampicillin at 50 µg/ml. Herein I will describe the fate of one of these cultures. The culture was

grown at 37°C in a shaking incubator, until it had reached an OD_{600} of 0.6 (approx. 2 to 3 hours). At this point an 100 µl sample (pre-induction) was taken for SDS-PAGE analysis, and stored at 4°C. Protein production was induced with IPTG (1M stock), which was added to the main culture to a final concentration of 4 mM. The culture was grown at 30°C for a further 2 hours. The reduced temperature was found to produce larger amounts of more active protein. After induction the cells were pelleted by centrifugation in 1L bottles at 3K for 10 minutes. The cells were resuspended in 50 ml of PBS, pH 7 with 1 protease inhibitor cocktail tablet (Sigma) pre-dissolved or with 4 mM PMSF. Another 100 µl sample (post-induction) was taken for SDS-PAGE analysis and stored at 4°C. The resuspension was then sonicated on ice with a large probe, for 30 seconds. Each resuspension was in turn sonicated between 4 and 5 times. A third 100 µl sample (post-sonication) was taken. The sonicated cells were then pelleted by centrifugation at 8K for 15 minutes. Supernatant and pellet were separated and an 100 µl sample of both was taken and stored at 4°C. The supernatant containing the soluble protein was removed and stored at –20°C. The supernatant and pellet were subsequently subjected to a series of purifications to remove it from all contaminating proteins.

2.2.16 Fast Method to test IPTG induced cell cultures for effective protein expression.

This method was devised to test IPTG induced bacterial cell cultures for efficient protein production before proceeding to full-scale purification of the protein batch. This is described for GST-fused proteins but can also be applied to other binding matrixes. A 40 µl sample was taken of glutathione sepharose 4B (Amersham-Pharmacia), the GST-binding material for protein purification. The suspension contains a mixture of glutathione beads suspended in ethanol. The beads were washed three times by spinning the suspension down slowly in an eppendorf tube for 5minutes at 5K, removing the ethanol supernatant and resuspending the beads in 40 µl of 1X PBS. Once the beads were washed, a 0.5 ml sample of the cell lysate supernatant was added to the beads, mixed gently and incubated at room temperature for 30 minutes. The beads were centrifuged at 5K for 5 minutes and the supernatant kept for analysis. The beads were then washed 3 more times with 40 µl of 1X PBS as

before but with each wash the supernatant was removed and kept for analysis. To each sample taken, a 20 µl of 2X SDS sample buffer was added. The 40 µl sample of pelleted beads was also taken for SDS-PAGE analysis. The samples were taken and run on a 10% SDS-PAGE gel to determine that the protein was (a) expressed and (b) correctly binding to the glutathione beads (data not shown). When the protein had been expressed to appreciable levels and was appearing to bind and elute correctly, the cell lysate supernatant was subjected to further purifications.

2.2.17 GST purification using Glutathione Sepharose 4B beads

Glutathione sepharose 4B beads (Pharmacia) were used to purify GST tagged proteins. Disposable 10 ml columns (BIO-RAD) were packed with 2 to 3 cm of glutathione sepharose 4B beads (GST-beads). The gel was washed through with 5-10 bed volumes of 1X PBS to remove ethanol and then equilibrated with 3 bed volumes of 1X PBS. The protein containing supernatant was filtered through a 0.45 µm filter before being applied to the column. In cases where the sample is still too concentrated, it was diluted in PBS + 1% Triton X-100. The protein extract was applied to the column and a sample kept back of the supernatant for testing. The eluent was collected and kept for analysis. The column was washed through with 10 bed volumes of 1X PBS to remove bacterial proteins from the applied lysate. The bound protein was eluted with at least 5 bed volumes of elution buffer. The elution buffer contained 50 mM to a maximum of 100 mM glutathione in 50 mM Tris-HCl, pH 8.0. The eluted protein was collected as 1 ml fractions. Samples of the lysate, eluent and fractions collected were all taken and run on a 12% SDS-PAGE gel. The GST-tagged protein was used in some initial Bandshift experiments (data not shown) and was later shelved in preference of a histidine tagged version of the MBD protein.

2.2.18 Expression of pET6hMBD; a histidine tagged MBD protein.

BL21(DE3)pLysS cells transformed with pET6hMBD; a histidine tagged protein expression vector containing the MBD domain of rat MeCP2 were used to inoculate a 10 ml culture of LB-broth and 100 µg of ampicillin and grown overnight at 37°C. The following day the culture was used to inoculate 500 ml of TB broth containing ampicillin @ 100 µg/ml. The culture was grown to an OD_{600} of 0.4-0.6 at which

point they were induced to express protein with IPTG to a final concentration of 0.4 mM. The cultures were grown for a further 2 hours after which point they were put in 1 L containers and pelleted by centrifugation for 10 minutes at 3.5K The cells were then resuspended in 15 ml of Buffer B with 1M urea and then sonicated three times each for 2 minutes each, on ice. The cell debris was pelleted by centrifugation for 35 minutes at 13,000 rpm using a JA20 rotor, after removing an 100 µl sample (pre-spin) for SDS-PAGE analysis. After the spin, another 100µl sample was removed from both the supernatant (post-spin-sup) and pellet (post-spin pellet). The pellet was then dissolved in 10 ml buffer B (+ 2.5 M urea + 0.8 M NaCl), and a sample removed for SDS-PAGE gel analysis.

2.2.19 Dialysis of the dissolved pellets from pET6hMBD expression.

To obtain the maximum amount of expressed MBD protein possible, pellets containing insoluble MBD protein from the induction procedure, above, were dissolved in buffer B containing 2.5 M urea and 0.8 M NaCl. To maintain their solubility they were dialysed slowly to remove the urea. This was achieved as follows. The dissolved pellets were carefully transferred to sterilised dialysis bags and clipped with sterilised clips. The dialysis bags were then placed in a 1 L beaker containing a 4°C solution of 1000 ml Buffer B with 2 M urea and 0.6 M NaCl. The pellets were allowed to dialyse for 1 hour at 4°C, while the buffer was being mixed with a magnetic stirrer. After this the buffer was exchanged to Buffer B containing 1 M urea and 0.4 M NaCl. Again, the buffer was precooled at 4°C, and the pellets were allowed to dialyse in the cold room at 4°C for 1 hour. Again, the buffer was exchanged to a pre-cooled Buffer B, containing 0.5 M urea and 0.2 M NaCl, and the procedure followed as described. The final two buffers for dialysis contained Buffer B containing (0.2 M urea + 0.1 M NaCl) and Buffer B only, in that order. In summary, the buffer order was as follows: (a) Buffer B (+ 2.5 M urea + 0.8 M NaCl), (b) Buffer B (+2 M urea + 0.6 M NaCl), (c) Buffer B (+ 1 M urea + 0.4 M NaCl), (d) Buffer B (+0.5 M urea + 0.2 M NaCl), (e) Buffer B (+0.2 M urea + 0.1 M NaCl) and (f) Buffer B only. Post-dialysis a sample was taken for gel analysis. Finally, the collective dialysed pellets were centrifuged at 15K for 30 minutes and a sample taken from both supernatant and pellet for SDS-PAGE analysis.

2.2.20 *Purification of the MBD protein pellet using a Fractogel EMD-SO_3^- -650 column.*

The isoelectric point of the MBD protein domain from $MeCP_2$ was estimated by using a web-based program from SWIS-PROT at http://www.expasy.ch/sprot. The sequence of the MBD domain of MeCP2 was obtained from Meehan *et al.*, 1993. From this information a purification protocol was designed. The MBD contains 85 amino acids with an estimated Mr= 10,000 Da. The estimated pI is 9.7. It is therefore a cationic protein and capable of binding to an SO_3^-column. The fully dialysed pellet of MBD protein as described above, was suspended in 25 ml of Buffer B. Having firstly retained an 100 µl sample for SDS-PAGE analysis later, the protein sample is applied to a 20 ml BIO-RAD disposable column containing 5 ml of Fractogel EMD-SO_3-cation exchanger that has been washed through several times with Buffer B. The flow through was collected and the column washed 3 times with Buffer B. To elute the bound protein the cation exchange gel was challenged with increasing concentrations of NaCl, effectively competing with bound protein for binding sites in the gel matrix. Buffer B containing concentrations of 0.4 M, 0.5 M, 0.6 M, 0.7 M, 0.8 M, 0.9 M and 1.0 M were in turn passed through the column in volumes of 1.5 ml. The fractions were collected and an 100 ml sample of each removed for subsequent gel analysis. The fractions containing a large quantity of the potential MBD protein were pooled. However, these fractions also contained a large amount of other SO_3^- binding proteins. Further purifications were required to remove excess protein.

2.2.21 *Purification of the Fractogel purified histidine tagged MBD protein over a Ni^{2+}charged affinity column.*

Fractions of protein collected from elution with NaCl in the range 0.5 M to 1M, described above were pooled. A Ni^{2+} charged column was set up by carefully adding 2ml of well-mixed Chelating Sepharose Fast Flow (Pharmacia) to a disposable 20 ml BIO-RAD purification column. The gel was allowed to settle and was subsequently washed through with 10 gel volumes of ddH_20. A 250 mM stock of $NiCl_2$ was prepared to a final volume of 50 ml in 50 mM Sodium acetate, pH 6. Three times the

gel volume of $NiCl_2$ solution was added to the column. As it flowed through the column the gel matrix shifted from white to green. The excess was washed through with several more gel volumes of ddH_20. When no more green eluent was removed from the column, the matrix was assumed to be at maximum retention of Ni2+ ions. The Ni2+ charged Sepharose was then equilibrated with 3 gel volumes of Buffer B. A 0.5 M stock of L-Histidine was made up in ddH_20. This was used to make a diluton series of different concentrations of L-histidine to effectively compete with the bound protein for Nickel ions. From this stock, L-histidine solutions were made in Buffer B to a final volume of 50 ml in the range 2 mM to 80 mM. An 100 µl sample of the pooled protein fractions was retained for analysis and the rest was applied to the column. The solution was allowed to drip slowly through and the flow through collected on ice. The bound histidine tagged MBD protein was eluted with increasing concentrations of L-histidine in Buffer B as 1 ml fractions. The experiment was repeated several times until the optimal concentration of L-histidine was determined for both maximum purity and maximum concentration of protein. An 100 µl sample of each fraction was taken for gel analysis and the rest of the protein fractions were snap frozen in liquid N_2 and stored at $-80^{o}C$. The collective samples from each fraction were run on a 10% SDS-PAGE gel.

2.2.22 Purification of supernatant fraction from pET6hMBD expression.

As described for pET6hMBD expression above, post IPTG-induced expression, bacterial cells were sonicated, centrifuged and the expressed protein was split between a soluble supernatant fraction and an insoluble pellet fraction. The pellet fraction was treated as described above, with urea solubilisation and subsequent column purifications over a cation exchange column and a Ni^{2+} charged column. The supernatant fraction was purified similarly, except that the dialysis as described was not required as the supernatant was already in a solution of Buffer B. Therein, the procedure of purification was performed on the supernatant fraction as it was for the pellet. Finally, purified fractions of the original supernatant from the nickel column were combined with solubilised and purified fractions from the pellet, reapplied to a nickel charged column and eluted to obtain a higher concentration of purified protein.

2.2.23 Annealing of complementary ssDNA oligomers

Annealing of DNA is carried out for substrates for Bandshift analysis and for substrates for the HIV-1 integrase assay. One radiolabelled (if required) primer is annealed by incubating it with a 1.2 M excess of the complementary strand in an 100 µl solution of 50 mM Tris, pH 7.5 and 10 mM $MgCl_2$. The mixture is placed in a 85°C heating block for 5 minutes. The block is removed and placed on a metal stand at room temperature to allow the block to cool gradually. This ensures correct template annealing and the elimination of any secondary structure anomalies.

2.2.24 Bandshift analysis using bacterial cell purified proteins

A selection of proteins were purified by virtue of having either a GST-affinity tag or a histidine affinity tag. Purified proteins obtained in this manner included (1) GST-MBD and HIS-MBD, the MBD domain of MeCP2 fused to a glutathione sepharose (GST) or a histidine (HIS) tag, respectively. (2) GST-MBD1, in collaboration with Dr. Irina Stancheva (3) GST-Kaiso-ZF, the Kaiso zinc fingers domain only, in collaboration with Dr. Alexey Rusov. However, purifications of Kaiso did not prove to be active and therefore bandshifts with this protein are not shown in this thesis. Subsequent purifications may be used by others in the laboratory. Highly purified fractions of each protein, as determined by Coomassie staining of SDS-PAGE gels were chosen for bandshift analysis. Protein concentration of each purification was estimated by using a BCA-200 protein assay kit (PIERCE) according to manufacturer's instructions and/or by spectrophotometry.

A variety of DNA oligomers were chosen for bandshift analysis. These were chosen based on published articles on the crystal structure of various dsDNA fragments in the literature. Table 1.2 provides a summary of that information and table 5.1 shows the oligomers that were synthesised. All oligos were synthesised by MWG Biotech. One of the ssDNA oligomers was 5'-labelled with ^{32}P-γ-ATP. The complementary strand was added in excess and annealed.

Bandshifts were performed in 1X Bandshift Buffer with increasing concentrations of purified protein. For each sample, approximately 10 ng (10 µl) of 5' labelled DNA is incubated with increasing concentrations of protein in 1X

Bandshift Buffer. For competitive bandshifts 1μg of DNA is added to each lane bar the control lanes. The volume is made to 50 μl with ddH_2O. Samples are incubated on ice for 30 minutes to allow protein binding to occur. A 5 % PAGE gel made with 0.5 X TBE is pre-run in the cold room. From each reaction 40 μl is taken to which 4 μl of glycerol is added. Normal 6X loading buffer interferes with protein-DNA interactions. Gels are run between 100 and 200 V in 0.5 X TBE. They are then dried down at 80°C for 40 minutes and exposed to either autoradiographic film or a Fugifilm PhosphorImager screen.

2.2.25 Protocol for Benzonase Digestions of Radiolabelled DNA

Substrate DNA for the Benzonase assay can be obtained from a variety of different sources. Plasmid DNA was generally prepared by midipreparation. DNA fragments generated by PCR were PAGE purified to remove primers and incomplete PCR products and cleaned up with ethanol precipitation followed by washing with 80 % ethanol. Fragments prepared by restriction enzyme digests of plasmid DNA were agarose or PAGE gel purified and the DNA extracted by cutting the relevant bands from the gel. Slices were placed in single eppendorf tubes along with 1 ml of Gel Extraction Buffer. The slices were left overnight on a rotating wheel in the cold room. The following day the supernatant was removed from the gel slices and was phenol-chloroform extracted, ethanol precipitated and resuspended in 50 μl TE buffer. Once DNA created by either PCR or plasmid digests is purified it is ready to be 5'-end labelled with ^{32}P-γ-ATP. One of the 5'-end must then be removed by restriction digests and the fragment is then ready to be digested by Benzonase. More details are provided in Table 2.5, which outlines the basic protocol used.

2.2.26 Protocol for DNAaseI Digestions of Radiolabelled DNA

The protocol for DNaseI analysis of DNA fragments is the same as that described above for Benzonase analysis. There are distinct differences in the DNaseI protocol

Lane No:	LANE 1	LANE 2	LANE 3	LANE 4	LANE 5	LANE 6	LANE 7	LANE 8
DNA	10	10	10	10	10	10	10	10
Carrier	1	1	1	1	1	1	1	1
10X Buffer	5	5	5	5	5	5	5	5
Benzonase	0	4	1.5	6.5	1	2.5	5	1
ddH20	34	30	32.5	27.5	33	31.5	29	33
Total	50	50	50	50	50	50	50	50
Benzonase Dilution Factor:		1/1000	1/100	1/100	1/10	1/10	1/10	1/1

Table 2.5 The Benzonase Assay. Typical parameters for the digestion of DNA fragments by the Benzonase Assay. All numbers refer to microlitres of solution. DNA is radiolabelled and typically between 100 ng to 1 μg DNA. Carrier DNA can be plasmid DNA minipreps or salmon sperm or *E. coli* DNA. Any of these have no bearing on the results of the digest as long as the carrier is unlabelled DNA. Carrier is typically made to a 1 μg/ml suspension in TE (10:0.1) mM or in ddH_20. 10 Buffer can vary between standard DNaseI reaction buffer to Buffer Y+ (MBI) or Buffer O+ (MBI). Benzonase operates well between 10 and 50 mM Tris.Cl solution and requires the presence of a divalent cation, typically Magnesium. The reactions are set up without Benzonase and each reaction is started by the addition of Benzonase. The dilution factor is based on units of Benzonase per μl of DNA. The reaction is timed for 5 minutes and stopped by the addition of 1ml of 0.5 M EDTA, 150 ml ddH_20 and 200 ml of phenol. All reactions are phenol-chloroform extracted and the DNA precipitated with Ethanol overnight in the presence of additional carrier DNA. DNA samples are spun down and resuspended in 1X Sequencing Gel Buffer. Samples are heat denatured and rapidly cooled prior to loading of denaturing PAGE/Urea gels.

Lane No :	1	2	3	4	5
DNA	10	10	10	10	10
CARRIER	1	1	1	1	1
$MgCl_{l2}$ (50mM)	5	5	5	5	5
DNaseI(10mg/ml)	0	1 x 1/64	1 x 1/32	1 x 1/16	1 x 1/8
H_20	39	38	38	38	38
TOTAL	50	50	50	50	50

Table 2.6 DNaseI analysis of DNA fragments. These are the parameters and reaction conditions used for DNaseI digestion analysis as shown in figure 3.6. DNaseI was obtained from ABgene as a freeze dried powder. It was dissolved in 0.1 M Tris, pH 7.51 as recommended and diluted to 10 mg/ml. It can be used between 1 - 100 μg/ml with 5 mM $MgCl_2$, pH 7.0 - 8.5. Further serial dilutions were made for the purposes of the assay and are shown in the table above. Reactions were set up in separate tubes with all of the componants except DNaseI. All values in the table are in μl of solution. Typically 10 μl of 5' end labelled DNA is used if the incorporation was deemed sufficient. This would represent about 200 ng to 1 μg per lane. Carrier DNA is 1 μl of 1 μg/μl of either Salmon Sperm DNA, *E. coli* DNA or suitable purified plasmid DNA. Reactions are timed for 5 minutes and stopped by the addition of 2 μl of 0.5 M EDTA. The reactions are phenol-chloroform extracted and precipitated with ethanol and extra carrier DNA overnight at -20°C. The samples are spun down and resuspended in sequencing gel loading buffer. Samples are run on an 8 % denaturing PAGE/Urea gel and autoradiographed.

	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7	Lane 8
DNA	10	10	10	10	10	10	10	10
5X Buffer bandshift buffer	20	20	20	20	20	20	20	20
CARRIER 1μg/μl	1	1	1	1	1	1	1	1
MBD 0.08 μg/μl	0	2	4	8	10	15	20	30
$MgCl_2$ (25 mM)	20	20	20	20	20	20	20	20
DNaseI (0.05mg/ml)	0	1	1	1	1	1	1	1
H_20	48	46	44	40	38	33	28	18
TOTAL	100	100	100	100	100	100	100	100

Table 2.7 DNaseI Footprinting Protocol. This is a typical protocol for DNaseI Footprinting and is the scheme used in figure 5.11. All values are given in terms of μl of solution. The 5X Bandshift Buffer consists of 100 mM Hepes, 5 mM EDTA, 15 mM MgCl2, 50 mM B-mercaptoethanol, 50 % glycerol and 0.5 % Triton X-100. Carrier DNA must be a non-methylated DNA sequence. Therefore only *E. coli* DNA or another generally non-CpG methylated DNA can be used as carrier. Salmon Sperm DNA is not suitable as carrier for this process. MBD concentrations were estimated by a BCA-200 protein assay kit (PIERCE) and by spectrophotometry. Reactions were set up without DNaseI (Pharmacia). Reactions were started by the addition of DNaseI and timed for 5 minutes. After 5 minutes the reaction was stopped by the addition of 2 μl of 0.5 M EDTA and placed on ice. Samples were phenol-chloroform extracted and precipitated with ethanol overnight at -20°C. Samples were spun down, dried and resuspended in 1X Sequencing gel loading buffer. They were then applied to a denaturing PAGE/Urea gel for separation.

after DNA fragment preparation. An example of the basic protocol with more details pertaining the DNaseI assay specifically is provided in Table 2.6.

2.2.27 Protocol for Footprinting Analysis of Radiolabelled DNA

The protocol for Footprinting analysis on DNA fragments is the same as that described above for Benzonase analysis. There are distinct differences in the Footprinting protocol after DNA fragment preparation. An example of the basic protocol and information specific to the Footprinting procedure used in this thesis is provided in Table 2.7.

2.2.28 Western Blotting Procedure for histidine tagged or GST tagged proteins.

An SDS-PAGE gel is run as normal on a small BIORAD protein gel system at 200 V for 40-60 minutes. The gel is soaked in 1X Transfer Buffer for 10 minutes to leach out the SDS. At the same time, one gel-sized sheet of nitrocellulose and two of 3MM paper are also soaked in the same buffer. The BIO-RAD transfer apparatus is set up according to the manufacturer's instructions. The gel is then layered on one side with the nitrocellulose sheet and on both sides with 3MM paper. Outside of the 3MM paper, a sponge is placed. The sandwich is placed in a holder and the gel transferred at 200 mA for 1 hour with a stirrer and cooling bar. Pre-stained markers help to monitor the transfer. After transfer the gel is washed with H_2O, stained with a 1/20 dilution of stock ponseau S to identify bands. Molecular weight bands can be marked with a pencil. The filter is destained with ddH_2O and the filter blocked for 20 minutes with PBS/2% dried non-fat milk/0.1% azide. The filter is then incubated from 1-4 hours to overnight with a 1/1000 dilution of antibody in TBS/1% FCS. The blot is then washed three times for 15 minutes with TBS/0.5 % Tween 20. HRP conjugated anti-mouse rabbit IgG at a 1/3000 dilution in TBS/10 % FCS for 1 hour. The blot was again washed three times for 15 minutes with TBS/0.5 % Tween 20 and then finally with ddH_2O. Commercial ECL solutions were mixed according to instructions and applied to the filter. After 1 minute the filter was wrapped in clingfilm, placed in a cassette and autoradiographed to develop the fluorescent reaction of the antibody and conjugated substrate.

Chapter 3: Results (I)

3.1 Devising a method to detect the presence of A-DNA in CpG methylated DNA.

3.1.1 Introduction

The background of this thesis has been detailed in the introduction (chapter 1) and is concerned with CpG methylated DNA structure and its' relevance in the context of higher order chromatin structure and gene regulation. The aims of this thesis as described in section 1.7, were to devise a method or to use/modify an existing one, to investigate whether CpG methylated DNA can induce the A-form tertiary DNA structure in gene regulatory regions. The Abstract describes the course of action taken in this work. Prior to establishing Benzonase as the choice enzyme for this analysis, other methods were investigated. In this chapter I will introduce these methods and also describe in detail how the Benzonase enzyme was applied for use on CpG methylated DNA.

The first method tried for the detection of A-DNA was a retroviral integration based assay. The experimental data from this approach is detailed in the appendix (chapter 7). This assay proved to be a complicated method of choice. It is dependent on the action of the HIV-1 integrase protein. The integrase protein *in vivo,* directs the integration of the viral HIV-1 genome into host DNA. It does this by recognising and binding to long terminal repeats (LTRs) at the ends of the viral DNA and recombining them with nicked host DNA by the method of site-specific recombination. The integrase assay developed by Chow *et al.*, (review: Chow, 1997) uses *in vitro* expressed and purified HIV-1 integrase to integrate a substrate sequence of interest into a target plasmid. Figure 3.1 summarises the procedure. The substrate DNA is first constructed by PCR to contain a copy of the HIV-1 LTR at each end, which HIV-1 integrase can recognize and bind. It then integrates many copies of the substrate into the target plasmid. After integration the target plasmid should contain copies of the substrate DNA at the preferred sites of integration for HIV-1 integrase.

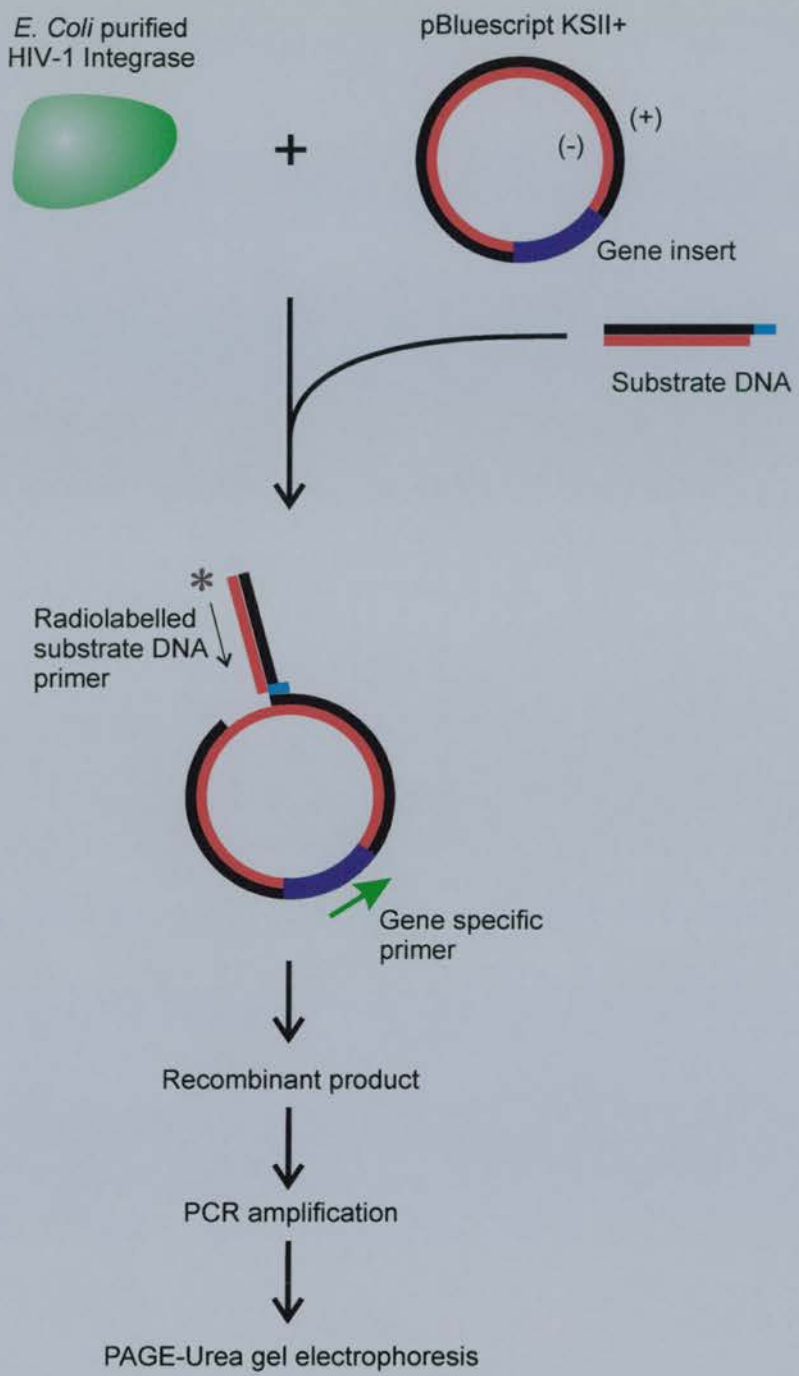


Figure 3.1 The Integrase Assay: An Outline. Schematic representation of the PCR based assay. The preprocessed substrate DNA has viral DNA recognition sites for HIV-1 integrase. Both protein and substrate combine and by the action of HIV-1 integrase, the substrate DNA is inserted into the target plasmid containing the gene of interest. The sites of integration are based on preferences of the HIV-1 integrase protein. These are usually regions of altered DNA structure; ie: bent DNA, nucleosomal DNA and CpG methylated DNA (Muller & Varmus, 1994; Pruss *et al.*, 1994a & 1994b; Kitamura *et al.*, 1992). These integration sites can be identified by PCR, as depicted, with a radiolabelled primer to the substrate DNA and a primer to a region in or 5' to the target gene of interest. Amplified products can be separated on a denaturing PAGE/Urea gel. This method can be used to identify the effect of CpG methylation on insertion site preference and therefore on DNA structure. (Adapted from Chow, 1997).

These include DNA with a wide major groove, unusual secondary structures, CpG methylated DNA and nucleosomal DNA (Kitamura *et al.,* 1992; Pruss *et al.,* 1994a & 1994b, Muller & Varmus, 1994). Detection of integration sites is via PCR with a radiolabeled forward primer specific to the target plasmid and an unlabelled reverse primer specific to the integrated substrate DNA. The resultant PCR fragments can then be run on a denaturing PAGE/urea gel. The position of each fragment on the gel should then correspond to the site of integration in the plasmid. Some sites will show preference over others.

It was hoped that CpG methylated DNA would provide hotspots for integration, with highly integrated regions showing pronounced structural differences from their unmethylated counterparts. The plan was to then investigate further the nature of these methylation dependent altered sequences. However, several problems were encountered as the assay depended on many experimental variables. These problems are discussed in the Appendix, Chapter 7. Additionally, the assay would only highlight DNA regions corresponding to preferred integration sites of unusual or an altered structure from typical B-form DNA. It does not tell us what they are and does not differentiate between small secondary structure alterations and larger conformational changes. It detects them all, as long as they provide a good binding site for HIV-1 integrase (Muller & Varmus, 1994). Finally, requests for a chemically synthesised chiral probe specific for A-DNA, which would help to isolate A-DNA from other structural alterations, was declined by the researchers concerned (Mei *et al.,* 1988). This approach was then abandoned.

It was subsequently found that research into the bacterium *Serratia marcescens* led to the discovery of an extracellular DNA/RNA non-specific endonuclease called benzon nuclease, now known commercially, as 'Benzonase' (BDH), (Meiss *et al.*, 1995). Investigations revealed that this enzyme has preferred sites of nuclease activity. In these experiments Meiss *et al.,* used PCR to generate specific DNA fragments from pBR322. The fragments were γ-^{32}P-ATP labeled at one end. The labeled fragments were digested with Benzonase and the fragments separated on a denaturing PAGE/urea gel. This work showed that regions of DNA known to contain regions that form A-DNA crystals, reacted strongly with the enzyme *in vitro* (Meiss *et al.*, 1999).

The above experiments did not involve the analysis of methylated sites, so we decided to try and use Benzonase for methylated DNA analysis. If Benzonase proved to be reactive with CpG methylated DNA, it would provide a more direct link to establishing sites of potential A-form DNA, than any retroviral integration assay could establish. It was also commercially available in highly purified and concentrated form as it is used in the large-scale removal of nucleic acids in pharmaceutical products. It would also be a novel use of this enzyme in methylation analysis, and perhaps tell us a little more about DNA structure and gene regulation. For these reasons, Benzonase was investigated as a potential tool for methylation analysis.

In this thesis, I have investigated the effects of Benzonase on DNA further. Initial questions included: Can Benzonase be used in a qualitative-quantitative assay to detect small (1-10 bp) regions of A-DNA? Can it be used to find methylation-dependent changes in DNA structure? Are these changes useful and informative in relation to gene regulation?

3.1.2 Results

3.1.2.1 Can the Benzonase Assay detect A-form DNA in CpG methylated DNA?

To investigate the effects of Benzonase on DNA, several initial experiments were performed to establish the correct concentration range to use such that there is on average 50 % digestion of the DNA fragments being analysed. This would allow visualization of a digestion pattern in the DNA, without extending to complete digestion of DNA. This experiment was repeated several times using the plasmid pBR322, and a large concentration range of Benzonase (data not shown). It was consistently found that Benzonase is a very potent nuclease. A dilution ratio of 1:10,1000, still provides strong nuclease activity on DNA. Having determined activity ratios of Benzonase to DNA, a pilot experiment was used to determine if any preference existed for CpG methylated DNA in terms of DNA digestion rates. Given that CpG methylated DNA can form A-DNA in certain sequence contexts (see introduction and table 1.2), it was reasonable to hypothesise that this might be reflected in Benzonase digestion rates on DNA. If CpG methylated DNA can form A-DNA, then perhaps it could be recognised and digested faster than unmethylated

DNA. The pBR322 plasmid was transformed into DH115 cells, midipreped, purified and split in half. DH115 cells were used as a host *E. coli* strain as they do not possess any intrinsic methylation activity, ensuring that methylation of adenine or cytosine has not already occurred on the sequence prior to analysis. One half of the DNA sample was treated with *Sss*I methylase which methylates the 5' position of cytosine in the context of double-stranded $d(CpG)_n$. It was then purified by phenol-chloroform extraction, ethanol precipitated and resuspended in 100 µl of ddH_2O. The complete methylation of the plasmid was tested by a methylation protection assay, an example of which is shown later in figure 3.12. The concentrations of both unmethylated and methylated pBR322 samples were estimated by spectroscopy at $A_{260/280}$ nm. The volume of each sample was adjusted to achieve the same DNA concentration in each sample. Both samples were divided into aliquots and subjected to a Benzonase digest for 5 minutes at increasing enzyme concentrations. The samples were purified and run on a 2 % agarose gel and visualized by staining with SYBR gold™ (Molecular Probes). As seen from figure 3.2, Benzonase appears to digest both unmethylated and methylated samples at similar reaction rates-with perhaps some increased sensitivity to methylated DNA. There was no indication from this experiment that CpG methylated DNA, (according to the hypothesis, having more regions forming A-DNA than the unmethylated sample), was digested appreciably faster by Benzonase.

However, it was also possible that CpG methylation alone was not enough to change the overall reaction dynamics, but perhaps the effects were localized and would only show altered cleavage dynamics at specific sites. It was decided to pursue this question using an assay based on the method of Meiss *et al.*, (1995, 1998, 1999). It was not clear at this stage whether the assay would be suitable for small stretches of A-DNA formation and whether it would be suited to methylation analysis. These aspects are addressed below.

3.1.2.2 Benzonase Analysis on the Amp^R gene of pBR322 : the effect of methylation

Some initial 'tester' DNA sequences were chosen for their known ability to form A-DNA by other methods. These were used to confirm that Benzonase would perform

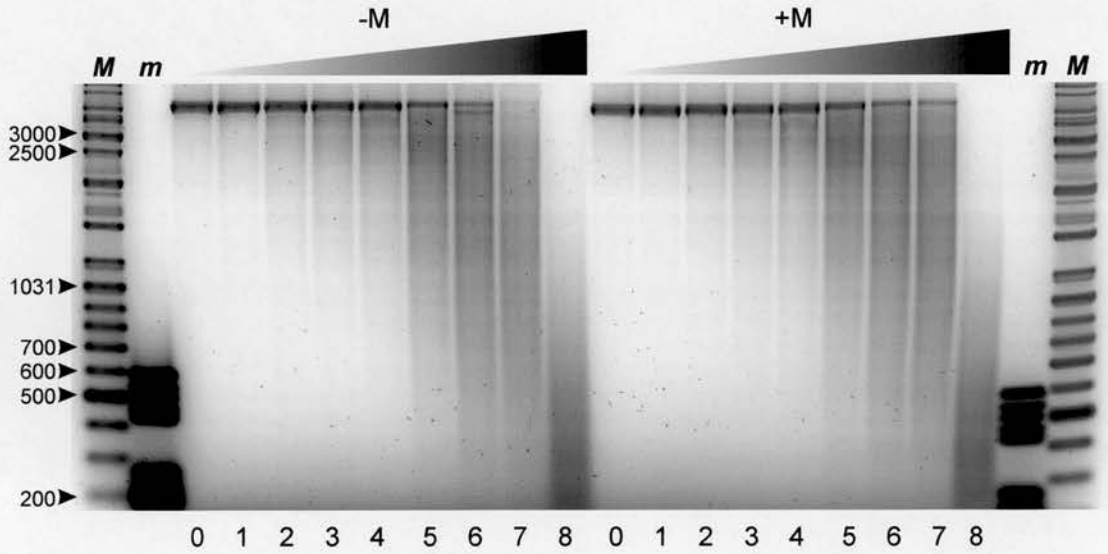


Figure 3.2 Benzonase digest of pBR322 plasmid. SYBR gold stained 2% agarose gel of benzonase digestion of *Sss*I methylated (+M) and unmethylated (-M) pBR322 plasmid. ***M***, Gene Ruler™ DNA Ladder Mix (MBI), ***m*** pBR322 digested with *Bsu*RI. Equal concentrations of pBR322 were subjected to a Benzonse digest for five minutes each, with increasing concentrations of Benzonase. The gel is stained with the sensitive nucleic acid fluorescent dye SYBR-gold (Molecular Probes), and scanned on a Fugifilm™ PhosphoImager.

in my hands, *in vitro*, as described by Meiss *et. al.*, (1988). The first of these tester sequences was previously shown to contain A-DNA (Mei *et al.*, 1998). In this study, a chiral probe, Tris(tetramethylphenanthroline)ruthenium(II) was synthesised and used as a tool to cleave A-DNA specifically, after activation with UV treatment, in conditions where little or no reactivity is seen with B-DNA. The same DNA sequence used by Mei *et al.*, was employed here to compare the reactivities of the chiral probe data to Benzonase digestion of the same sample.

The procedure adopted for the Benzonase assay both here and in subsequent analyses is summarised in figure 3.3. The protocol for digestion follows that described below for all Benzonase digestions unless otherwise stated. More detailed information is provided in Chapter 2. For Benzonase analysis, *Hind*III - *Ssp*I DNA fragments were prepared from pBR322 by PCR. DNA fragments can also be prepared by large-scale plasmid digestion. The full sequence of the fragment used is shown in figure 3.4 *i.*and *ii.* The PCR products were then split in half, one half of which was *Sss*I methylated. The extent of methylation was confirmed by a methylation protection assay (data not shown). Both unmethylated and methylated samples were Benzonase digested and the fragments run on a 10 % denaturing PAGE/urea gel (figure 3.5 *A*). The marker lane was prepared from unmethylated and labeled *Hind*III – *Ssp*I fragments which were digested individually with *Sss*I, *Hha*I, *Dde*I, *Hae*III, *Eco*RI and *Cla*I. An aliquot of each of the digested fragments was run in separate lanes on a 10 % PAGE gel to confirm the correct cleavage and sizes of the DNA (data not shown) These fragments were combined and run alongside the Benzonase digested fragments as marker DNA. As can be seen from figure 3.5 *A* the unmethylated pBR322 fragments shows hotspots of Benzonase activity throughout the sequence analysed.

As this fragment is the same as that used in the study by Mei *et al.*, (1988), I have compared the ability of Benzonase to cleave A-DNA with the pattern generated by the A-DNA specific Ruthenium complex over the entire fragment. This is provided in figure 3.4 *iii.* As can be seen from the two patterns, Benzonase reacts in the same way with the pBR322 fragment as does the Ruthenium complex. This confirms that Benzonase is indeed a useful tool for the detection of A-form DNA, given the specificity of the Ruthenium complex for A-DNA (Lim & Barton, 1998).

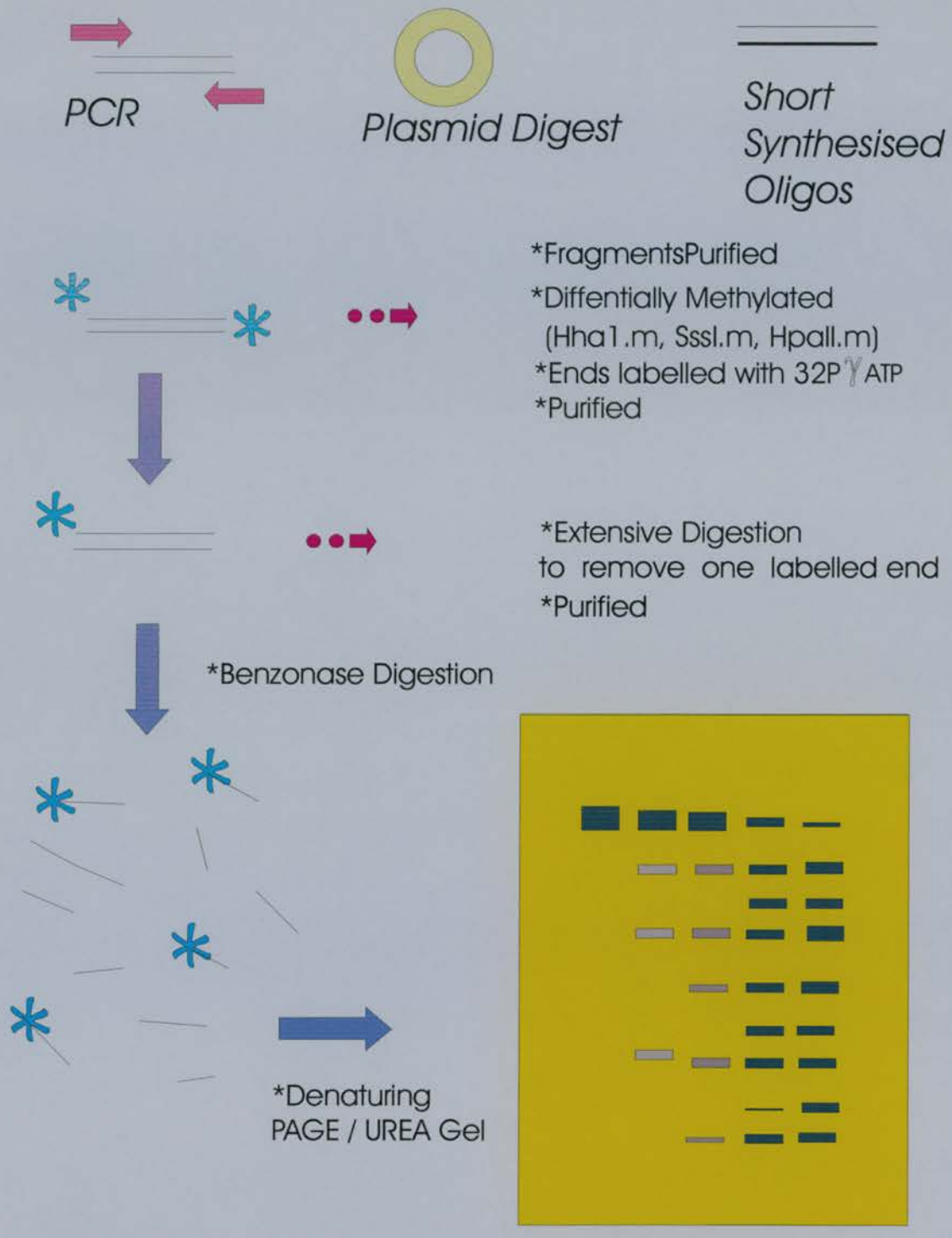
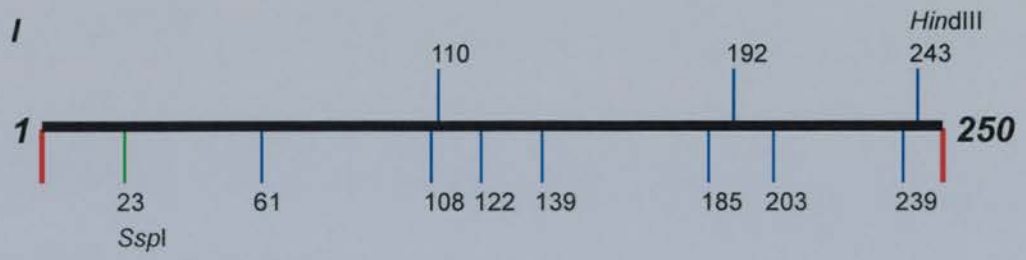


Figure 3.3 The Benzonase Assay: An Outline. Fragments for analysis can be prepared by large scale and purification of plasmid DNA containing the sequence of interest, then restriction digested to release the sequence, which can be separated and purified by agarose electrophoresis or PAGE. Sequences can also be generated by PCR or they can by artificially synthesised. Fragments can then be methylated with a variety of methylases. They are then γ-^{32}P-ATP labelled at both ends, with one end later removed by restriction digest. Labelled fragments are then subjected to Benzonase digestion and the resultant fragments purified and separated on a denaturing PAGE/Urea gel. The gel is then dried down and exposed to X-ray film and/or to a phophorimaging screen to identify the position of the radiolabeled digested fragments of DNA. The results are analysed on a Fugifilm PhospholImager machine.

i



ii

5'-ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT

CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA

TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCC

CCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATC

ATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT

TTCGTCTTCA AGAATTCTCATGTTTGACAGCTTATCATCG

ATAAGCTTTAATGCGGTAGTTTATCACA-3'
ATTACGCCATCAAATAGTGT-5'

Figure 3.4 The sequence of pBR322 used for Benzonase analysis. In ***i*** above is an illustration of the fragment from pBR322 used for Benzonase analysis. The exact positions of each CpG is indicated and relevant restriction sites. In ***ii*** above is the exact sequence used. In green are highlighted the PCR primers used to amplify this sequence. In yellow, d(AATATT) is an *Ssp*I site and d(AAGCTT) is a *Hind*III site. In blue, a *Hha*I site with two methylatable CpGs and the position where methylation enhanced cleavage is seen with Benzonase digestion. In red, other methylatable CpGs in the sequence that show reactivity with DNaseI but not with Benzonase (this thesis).Underlined is the sequence used by *Mei et al.,* 1988 to detect cleavage by a Ruthenium complex designed to target A-DNA. ***iii*** below shows the sequence used for detection of potential A-DNA using a chiral probe by Mei *et al.,* 1988 and is adapted from this paper. The region from pBR322 used by Mei *et al.,* is shorter than that used in this thesis, but overlaps with a central region of it as underlined above. The *Hha*I site was not incorporated into their study, nor was the effect of methylation on this sequence. The vertical lines indicate the degree of cleavage at each site (Mei *et al.,* 1988).

iii

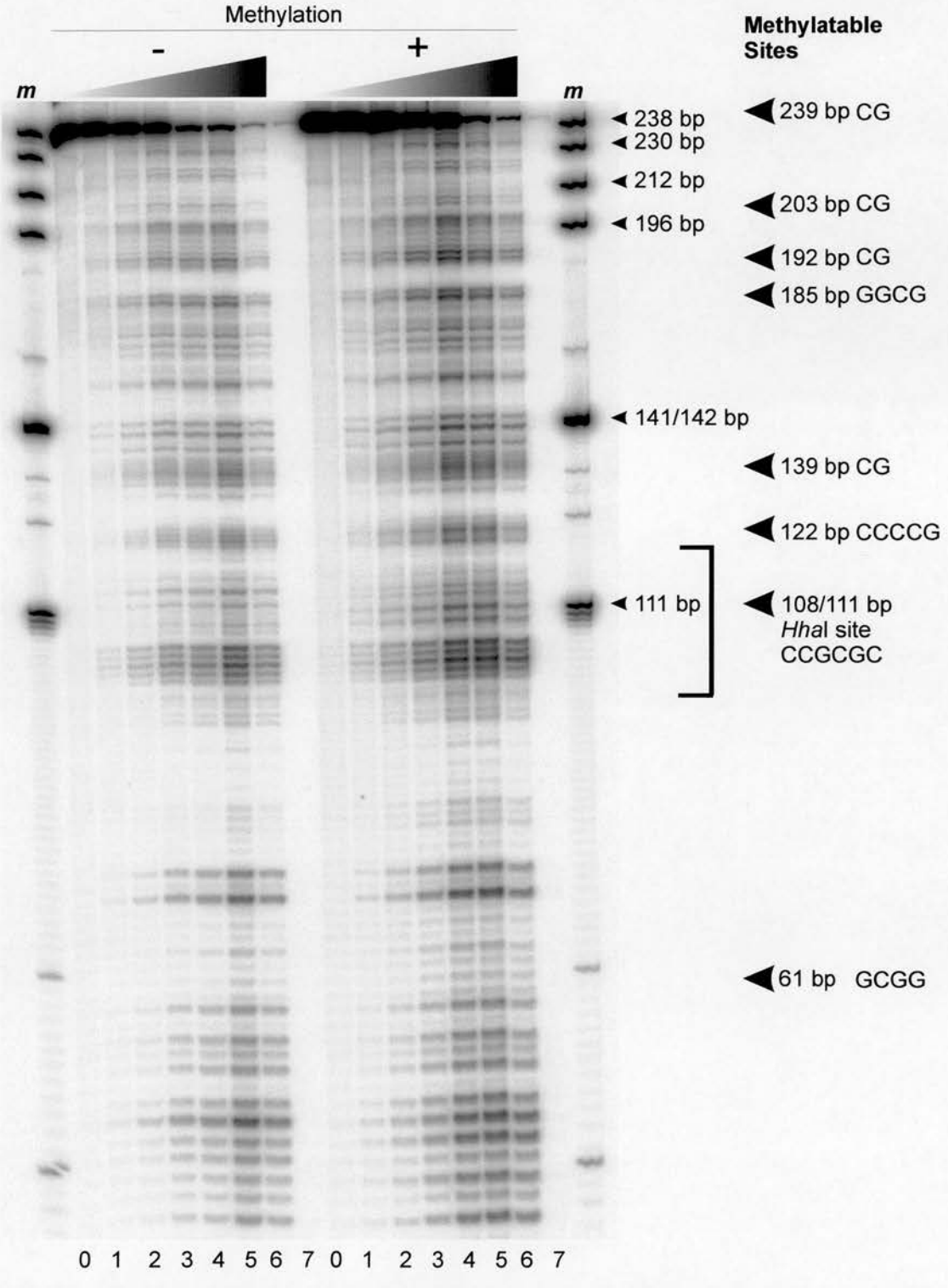


Figure 3.5 *A* Benzonase digestion of pBR322 fragments. Benzonase digestion of PCR generated pBR322 fragments in CpG methylated (+M) and unmethylated (-M) states. ***m***, marker lane generated by restriction digestion of unmethylated pBR322 samples, with *Cla*I, *Eco*RI, *Hae*III, *Dde*I, *Hha*I and *Ssp*I. The digested fragments were combined to create the DNA ladder shown here. The methylatable CpGs in the sequence are shown by large arrowheads along the margin. These are shown to indicate that the predominant methylation induced structural change in this sequence occurs at a *Hha*I site containing two methylatable CpGs adjacent to each other in an alternating $d(CG)_n$ string. The region bracketed is highlighted in figure 3.5 *B*.

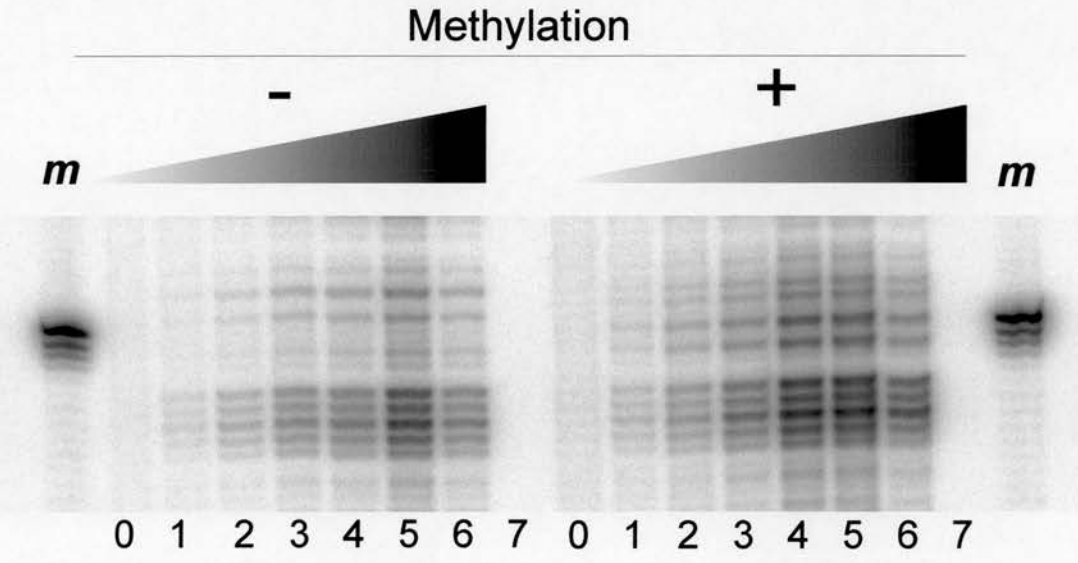


Figure 3.5 *B* Magnified view of Image 3.5 *A*. Benzonase Digest on pBR322. This enhanced view is to show clearly the methylation induced changes visible over a *Hha*I recognition site as detected by Benzonase on the methylated (+M) side of this gel. The marker is *Hha*I digested unmethylated pBR322 fragment which exactly pinpoints the region of structural change in the fragment.

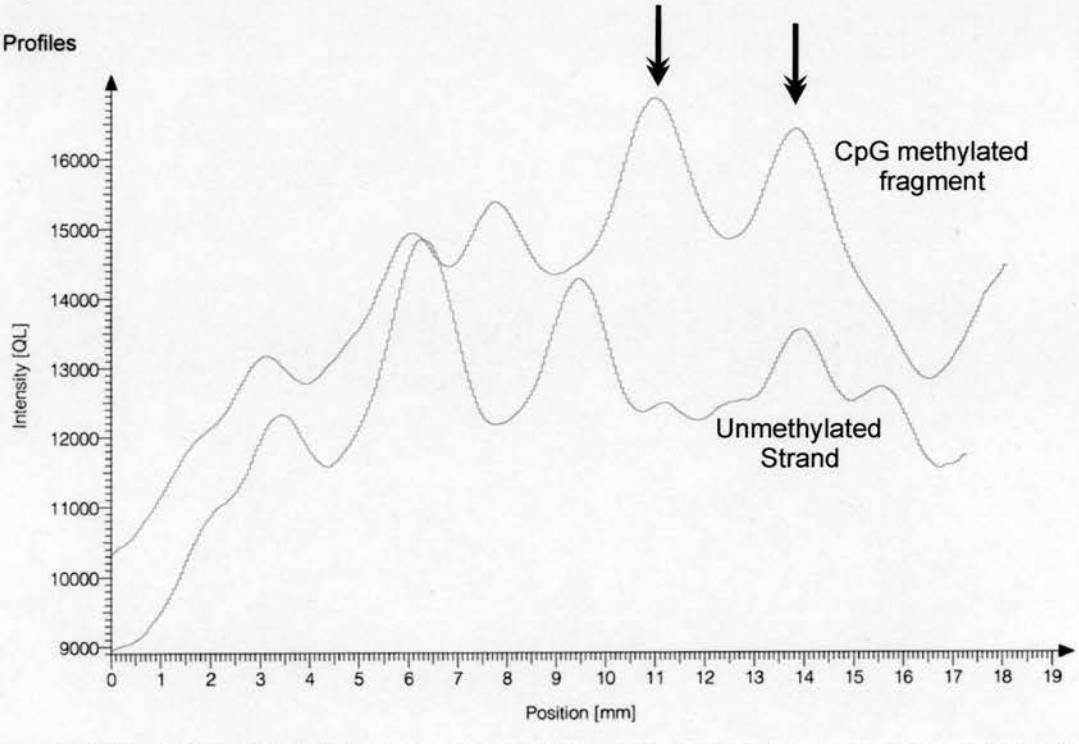


Figure 3.5 *C* Peak determination scan of figure 3.5 *A*. In this figure a region covering the methylation induced change at position 111 bp, bracketed in figure 3.5 A, has been scanned using Aida software in conjunction with a Fugifilm phosphoImager machine. The bracketed region from lane 5 for the unmethylated strand is compared to lane 5 for the methylated strand. Regions where CpG methylation has induced a distinct structural change, as detected by Benzonase, are indicated by an arrow. The intensity along the Y axis are arbitrary values assigned by the Aida software over the full scale of intensity values available in the data set. The X-axis refers to position within the gel over the region scanned.

The biggest revelation in this analysis is in the CpG methylated DNA. For the CpG methylated pBR322 fragment, there is a different banding pattern at one location only. The CpG methylation dependent change is highlighted and blown-up in figure 3.5 *B.* A peak determination analysis was performed on unmethylated and methylated regions covering the methylation induced structural change. This was performed using Aida™ software developed specifically for use with a Fugifilm PhosphoImaging machine. The results of this analysis are shown in figure 3.5 *C.*

There are ten CpG sites in the pBR322 fragment. Figure 3.4 *ii* shows the full sequence of the pBR322 fragment. The *Sss*I methylase enzyme has the ability to methylate all of the CpG sites highlighted in the text. From DNaseI analysis of the methylated sequence, this appears to be the case (see below). With this in mind, it is surprising that only one of these sites reacts strongly with Benzonase. When the DNA sequence surrounding each CpG was analysed, it was found that there was only one CpG that was contained in an alternating GC string. The rest, even though surrounded by cytosine and guanine residues, had neither sequence alternation nor adjacent methylatable sites-both features that would promote consistency and the formation of A-DNA. The highly reactive sequence at 108 bp and 110 bp, represents a *Hha*I cutting site. It has two adjacent CpGs, and alternation of sequence for 4 bp, (5'...CCGCGC...3'). In the entire sequence there are no other adjacent CpG doublets. It is reasonable to conclude that; from the alternating nature of this sequence, the fact that X-ray crystallographic studies have shown this type of methylated sequence to form A-DNA (Mooers *et al.*, 1995; Tippin *et al.*, 1997a & 1997b) and its reactivity with Benzonase, that this methylated sequence is capable of forming A-DNA. This is the first example of CpG methylated DNA showing sequence specific structural changes by this assay.

3.1.2.3 DNaseI analysis on the Amp^R gene of pBR322: the effect of methylation

DNaseI analysis was employed on the same pBR322 fragment as used above. There are several reasons for this analysis. DNaseI is also a strong endonuclease and has been employed as a useful tool in DNA structural analysis (see Sambrook, 1997 for applications). DNaseI however, has different site preferences to Benzonase. Unlike Benzonase, it reacts more strongly with the minor groove of DNA than the

major groove, but is much less regular in its cleavage patterns. As the 5'-position of cytosine lies in the major groove, any changes in cleavage rate by DNaseI must be due to DNA structural changes rather than altered hydrophobic contacts between the endonuclease and DNA (Fox, 1986). DNaseI shows altered cleavage patterns with unusual DNA secondary structures, such as kinks, bends, cruciform structures and methylated DNA. However, it cannot discriminate between them (Drew & Travers, 1984; Fox and Waring, 1984; Fox, 1986). DNaseI concentration was optimised in several experiments on both plasmid DNA and radiolabeled DNA to give approximately 50 % digestion rates (data not shown). A DNaseI assay was carried out on the sequence. The process for DNaseI analysis follows the illustration for Benzonase analysis shown in figure 3.3.with DNaseI in place of Benzonase. When DNaseI analysis is applied to the pBR322 fragments in unmethylated and methylated states (figure 3.6 *A*), it is clear that the DNaseI pattern is very different from the Benzonase pattern (figure 3.5 *A*). Figure 3.6 *A* shows that now, instead of one highly reactive methylated CpG region, there are now reactive sites at every CpG methylated site. The marker lane is the same mix of digested fragments as that used in 3.1.2.2 above. This assay confirms that: (a) DNaseI is capable of picking out every CpG in the methylated lanes of the digested pBR322 fragment. Therefore, the *Sss*I methylation reaction was highly successful. (b) Each methylated CpG contributes to an altered secondary structure in the DNA sequence that makes it more accessible to DNaseI activity, and (c) Benzonase is neither reacting randomly with methylation induced altered DNA secondary structure, nor regions that have a slightly widened DNA minor groove, as detected by DNaseI. It only prefers to digest particular regions of CpG methylated DNA.

There is only one prominent band in the DNaseI analysis that corresponds exactly with the Benzonase analysis. Next to the marker at 111 bp are at least two strongly reactive bands in the methylated lanes (figure 3.5 *B* and 3.6 *B*). At this position there are two adjacent CpG sites surrounded by alternating C and G residues, constituting a *Hha*I recognition site. It is this site that Benzonase reacts strongly with. As stated, this methylated sequence has been crystallised as A-form DNA (Mooers *et al.*, 1995; Tippin *et al.*, 1997a & 1997b).

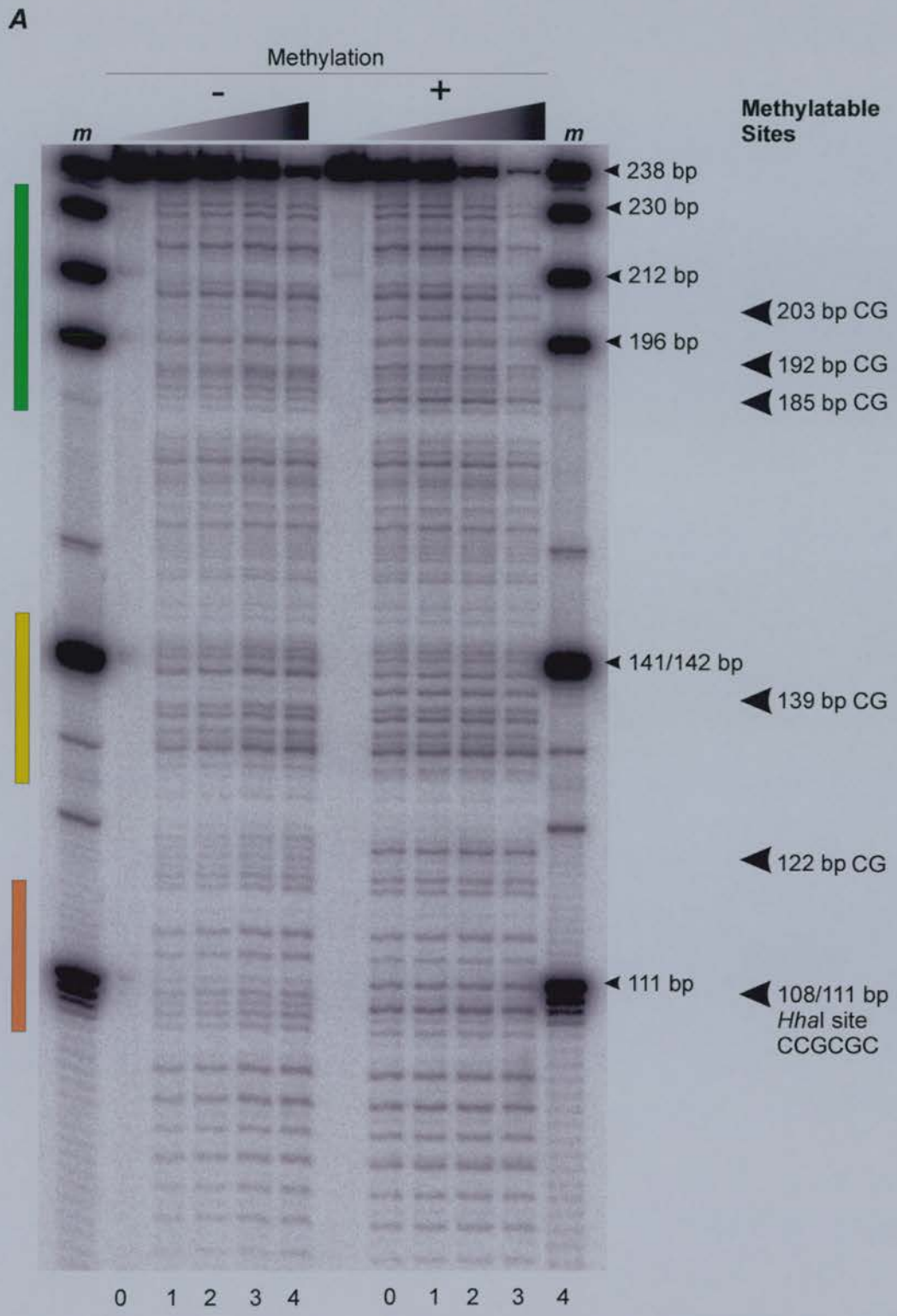


Figure 3.6 *A* DNaseI digestion of pBR322 fragments. ***A*** DNaseI digestion of PCR generated fragments from pBR322. These fragments correspond to the same fragments used for Benzonase analysis in Figure 3.5. ***m***, marker lane generated by restriction digestion of unmethylated pBR322 with *Cla*I, *Eco*RI, *Hae*III, *Dde*I, *Hha*I and *Ssp*I. Small arrowheads indicate the sizes of these combined markers. Large arrowheads indicate methylated sites in the DNA sequence that react strongly with DNaseI. The bars along the left hand side indicate regions analysed for peak determination scans, which are shown in figures 3.6 C (i, ii and iii).

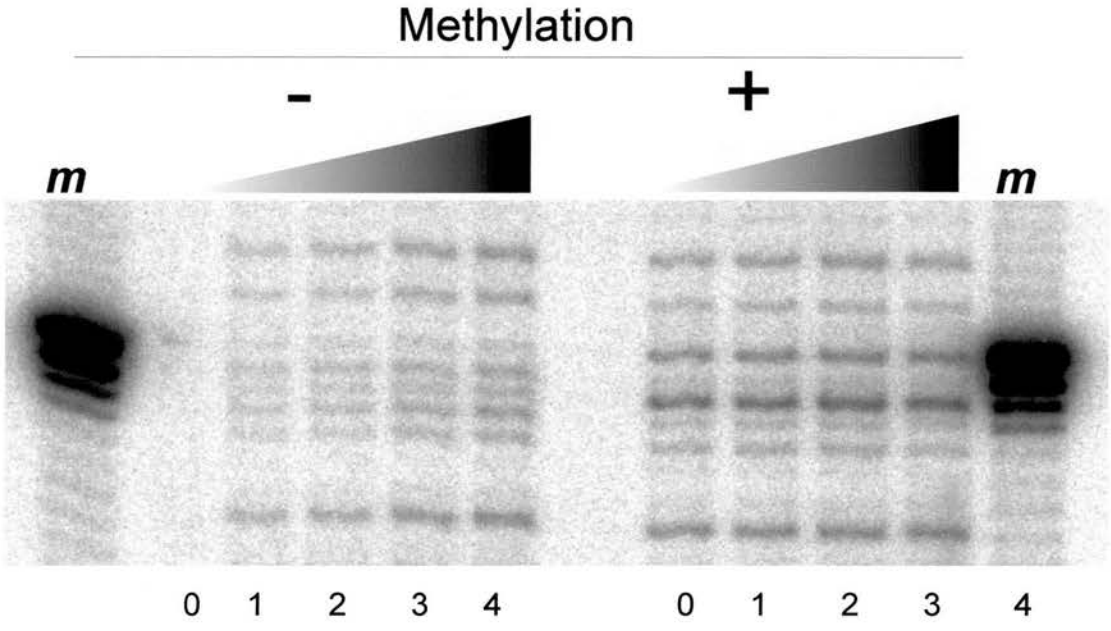


Figure 3.6 *B* Magnified view of Image 3.6 *A*. DNaseI analysis on pBR322. This enhanced view is to show the pattern of DNaseI digestion on unmethylated (-) and methylated (+) fragments of the pBR322 plasmid. This view is to demonstrate the enhanced cleavage at the CpG triplet, marked by a *Hha*I digest, lane m. Compare to Benzonase digest of same fragment above in figure 3.5 *B*.

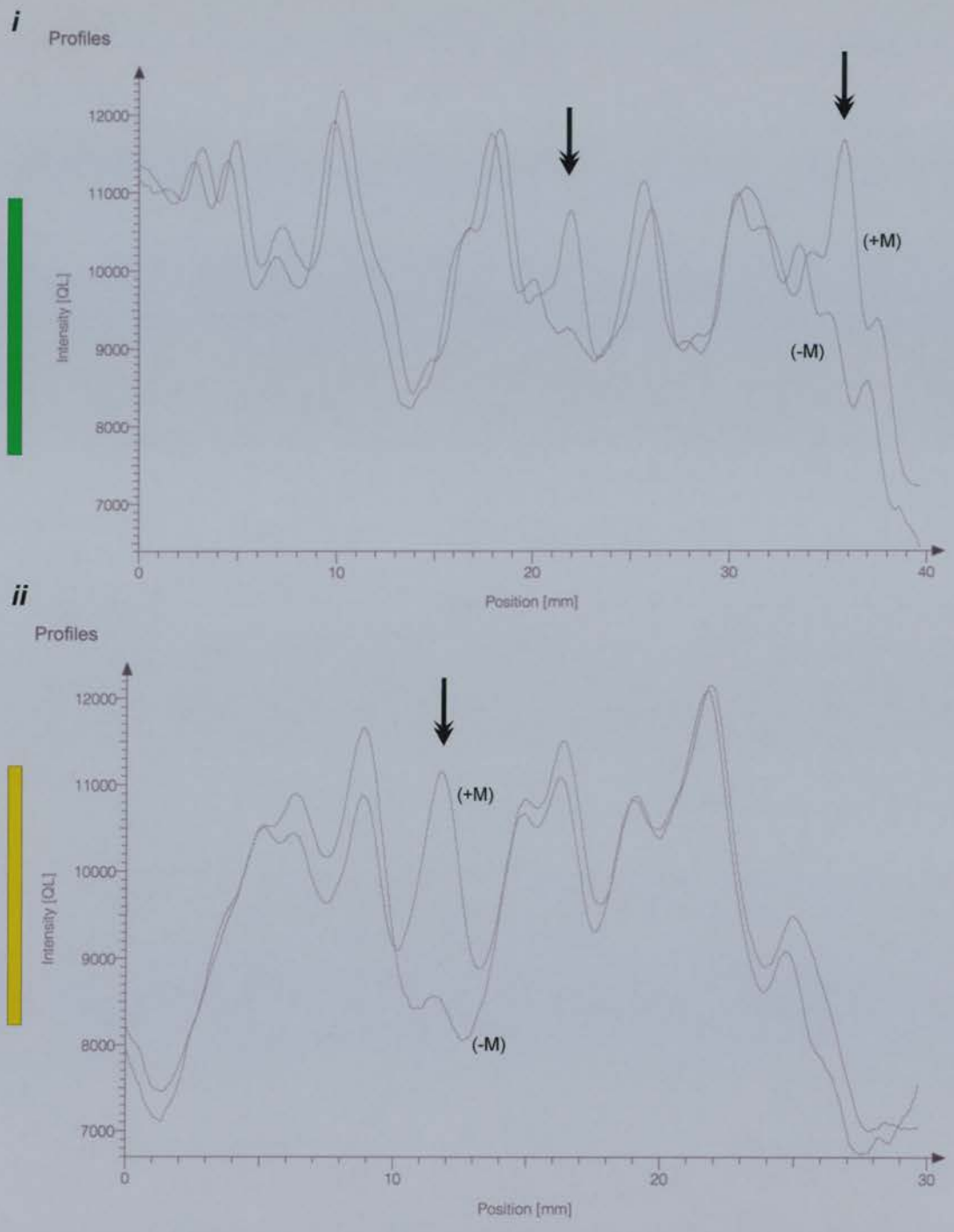


Figure 3.6 C Peak determination scan of selected regions of the DNaseI digestion analysis shown in figure 3.6 A. Selected regions were scanned using Aida software coupled to a Fugifilm PhosphoImager. The results of these scans are shown in figures ***i, ii*** above and ***iii*** on the following page. In all cases the regions scanned are colour coded on the left hand side of each figure. The same bars are shown in figure 3.6 A, indicating the regions selected from the gel. The intensity values are arbitrary values assigned by Aida software and refer to the complete range of intensity values available in the data set. Vertical arrows show regions of methylation enhanced cleavage by DNaseI. For the scans shown in figures ***i, ii*** and ***iii,*** lane 4 was chosen for the unmethylated strand and lane 3 was chosen for the methylated strand. These were judged to be the most similar in terms of overall intensity values by Aida software and therefore suitable for comparison. (+M) is the CpG methylated fragment. (-M) is the unmethylated fragment.

iii

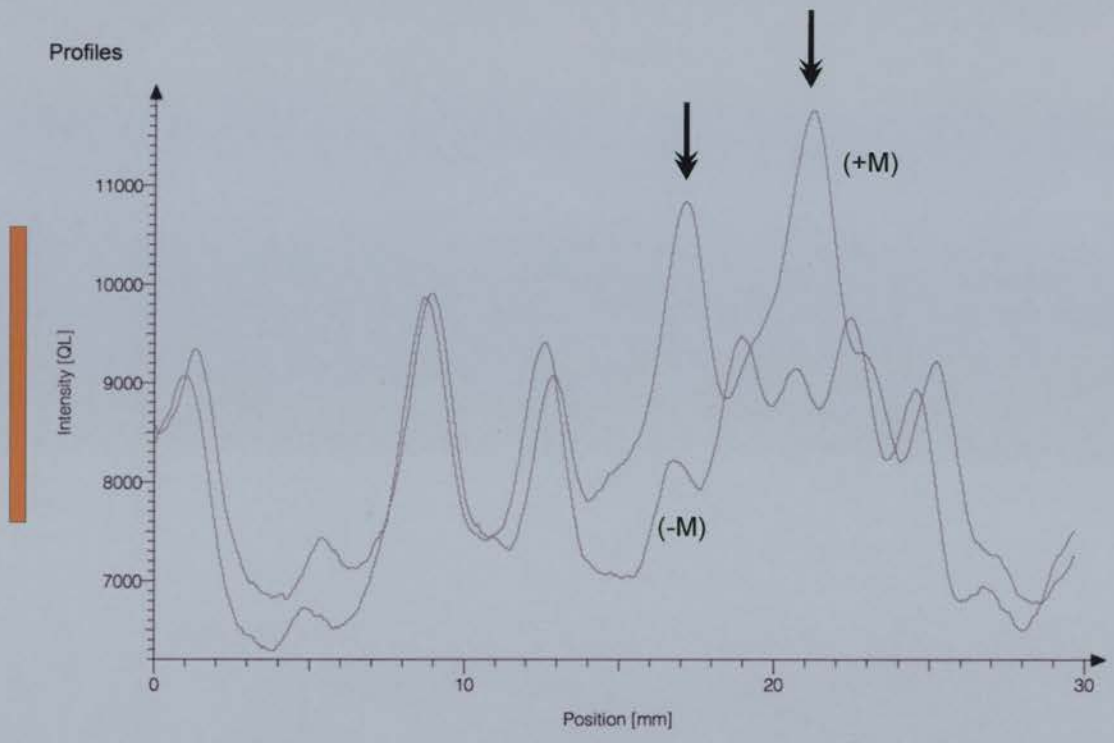


Figure 3.6 C Peak determination scan of selected regions of the DNaseI digestion analysis shown in figure 3.6 A. Selected regions were scanned using Aida software coupled to a Fugifilm PhospholmagerThe region scanned is colour coded in orange on the left hand side of each figure. The same bar is shown in figure 3.6 A, indicating the region selected from the gel. The intensity values are arbitrary values assigned by Aida software and refer to the complete range of intensity values available in the data set. Vertical arrows show regions of methylation enhanced cleavage by DNaseI. For the scans shown in figures ***i, ii*** and ***iii,*** lane 4 was chosen for the unmethylated strand and lane 3 was chosen for the methylated strand. These were judged to be the most similar in terms of overall intensity values by Aida software and therefore suitable for comparison. (+M) is the CpG methylated fragment. (-M) is the unmethylated fragment.

Peak-determination analysis was performed on the DNaseI digested fragments shown in figure 3.6 *A*. Unmethylated and methylated regions of figure 3.6 *A* were chosen for analysis with Aida software linked to a Fugifilm™ phosphoImaging machine. The results are shown in figure 3.6 *C* (i, ii and iii). Three regions of the DNaseI digestion analysis were selected that showed pronounced reactivity with DNaseI. Regions selected are clearly marked in each figure. The results clearly demonstrate the degree of alteration of the sequence in response to methylation at specific sites.

The fact that Benzonase shows high reactivity with this sequence only, while there are clearly other methylation induced structural alterations in the sequence, is a tribute to its' specificity. Of note here also is the fact that stable A-DNA intermediate structures have been crystallised. It has been shown that this transition is completely promoted by CpG methylation. Vargason *et al.,* (2000) have shown that the sequence $d(GGCGCC)_2$ forms B-DNA but the sequence $d(GGCGm^5C)_2$ forms E-DNA, a stable intermediate between A-and B-DNA forms. As this sequence also exists in the analysed fragments here, it is of note that it does not seem to be picked out by Benzonase. Perhaps Benzonase requires the complete transition to A-DNA before it can react readily enough to show heightened interaction at methylated CpGs? Interestingly, a stable intermediate can also be seen for the B-DNA transition in non-methylated DNA sequences (Ng *et al.,* 2000). Both studies illustrate the fluidity of DNA structure. It is not a rigid uniform structure, but dynamic in its changeability, a feature that is echoed in this thesis.

3.1.2.4 Benzonase analysis of a fragment of the pGEM9zf(-) plasmid vector

From the above studies it is clear that Benzonase can be used effectively for methylated DNA analysis. It can distinguish between small structural changes that are more readily picked out by DNaseI analysis and larger conformational changes in DNA. It is established that Benzonase can react in the same way with a pBR322 fragment as a chiral ruthenium complex that specifically recognizes A-DNA. To further confirm that Benzonase was behaving as hoped and expected, a sample of the original tester DNA from Dr. Pingoud was kindly donated to the project (Meiss *et al.,* 1999). The plasmid pGEM9zf(-) contains an 167 bp tester fragment which was amplified by PCR. Herein it is called 'Pingoud'. These fragments were purified, with

half the volume *Sss*I methylated. Both unmethylated and methylated samples were γ–^{32}P-ATP labeled at both ends with one of the labeled ends removed, in this case the *Sal*I end. The Pingoud fragment used in this analysis is shown in figure 3.7 with the CpGs highlighted. They were then subjected to Benzonase analysis. The results of this experiment are seen in figure 3.8 *A*. It is apparent that Benzonase is behaving as it does for Meiss *et al.*, 1999. The unmethylated lanes show the same hotspots for Benzonase activity as in the original study. However, it is also very apparent that there is a major difference now in the methylated fragments. As indicated in figure 3.8 *A*, there is a major shift from the unmethylated to methylated fragments. At this region the sequence contains a string of $(CG)_n$ dinucleotides that are unreactive to benzonase. However, when these dinucleotides are methylated they become highly reactive to Benzonase. A much longer stretch of nucleotides are affected by methylation here than in the previous analysis, which centred on a specific alteration. This region was selected for Peak determination analysis, the results of which are shown in figure 3.8 *B*. A similar experiment was performed using *Hha*I methylase (methylates the central cytosine in GCGC) instead of *Sss*I methylase. The pattern of digestion showed the same region of enhanced interaction with Benzonase confirming that it is alternating and methylated CG rich DNA that alters DNA structure (data not shown).

These initial experiments highlight both the usefulness of Benzonase in structural analysis of small CpG induced structural transitions and larger structural transitions. It also highlights the very dramatic changes that DNA is capable of, when CpG methylated. Such large structural alterations must have a strong regulatory role *in vivo*, but these aspects of gene regulation in eukaryotes are still poorly understood.

3.1.2.5 Refining the technique: The use of urea to promote specificity of Benzonase

Meiss *et al.*, (1995) investigated the effects of various reaction conditions on Benzonase activity and specificity. They demonstrated that the use of urea could enhance the effect of Benzonase at preferred cut sites. This aspect of Benzonase activity was reinvestigated here in an effort to refine the technique for A-DNA more

i

ii

5'-GCAGATTTAGGTGACACTATAGAATATGCATCACTAGTAAGCTTTGCT

CTAGAGGGCGCCCCAATTAAAAAACAATTGGTATACCCAATTCCAACGT

TGGCAATTCGCGCGCGCAATTGGGGGGCAATTGCCCGGGCGTCGACG

AGCTCCCTATAGTGAGTCGTATTA-3'

Figure 3.7 In *i* above is an illustration of the fragment from pGEMZf9(-) used for Benzonase analysis. This is the same fragment as that used by Meiss *et al.*, (1999). The position of each CpG is indicated in red. In ***ii*** above is the exact sequence used. In green are highlighted the PCR primers used to amplify this sequence. In yellow, d(TCTAGA) is an *Xba*I site and d(GTCGACGA) is a *Sal*I restriction site. Underlined are *Hha*I recognition sites and the positions where methylation enhanced cleavage is seen with Benzonase digestion (this thesis). In boxes are sequences that correspond to different types of DNA tertiary structure as deduced from several CD, NMR or crystallographic studies. $d(GGGCGCCC)_2$, A-type, (Eisenstein *et al.*, 1988); d(A)d(T)-tract, B-type, (Alexeev *et al.*, 1987); $d(GGTATACC)_2$, A-type, (Weston *et al.*, 1992; Zhou *et al.*,1987, 1988; Wang *et al.*,1987; Jamin *et al.*,1985); $d(CCAACGTTGG)_2$, B-type, (Prive *et al.*, 1991); $d(CGCGCGCG)_2$, B-type, (Butkus *et al.*, 1987); d(C)d(G)-tract, A-type, (Heinemann *et al.*, 1994); $d(GCCCGGGC)_2$, A-type, (Heinemann *et al.*, 1987, 1991). The 5'-CAATT-3' fragment shown in purple is used to separate the different structural motifs. The plasmid pGEMZf9(-) containing the above 167 bp insert was a gift from Dr. Meiss. (Adapted from Meiss *et al.*, 1999).

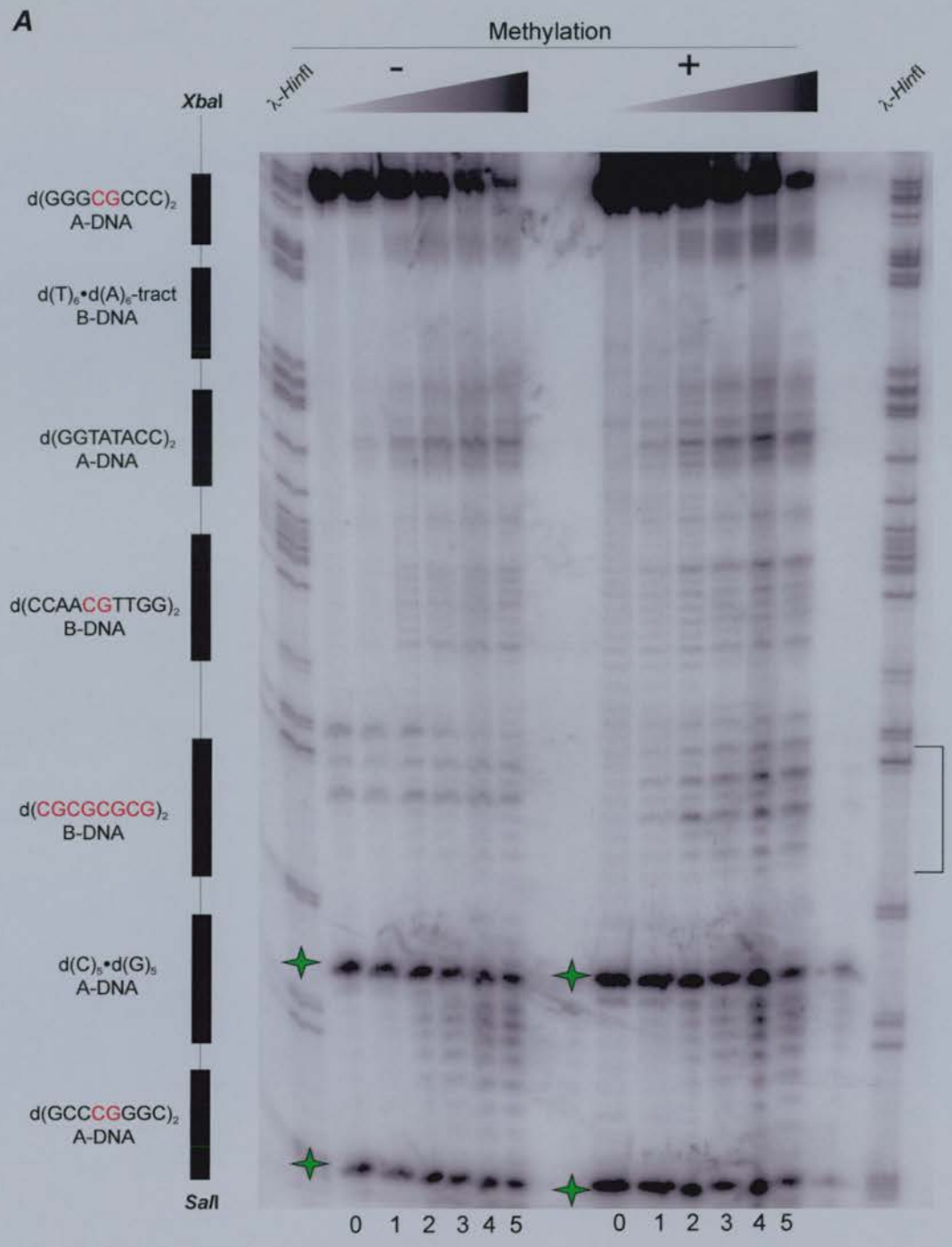


Figure 3.8 *A* Benzonase digest on pGEM9Zf(-) fragments. Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) PCR generated fragments from pGEM9Zf(-).λ-*Hin*fI, DNA digested with *Hin*fI. The fragments were labelled at one end and the fragments run on a 6 % PAGE / Urea gel. There is a repeat of this experiment shown in figure 3.10 which shows the methylation dependant changes in a higher percentage gel and is therefore clearer. Alongside are the regions of the 167 bp fragment used. Star symbols indicate anomalous bands.

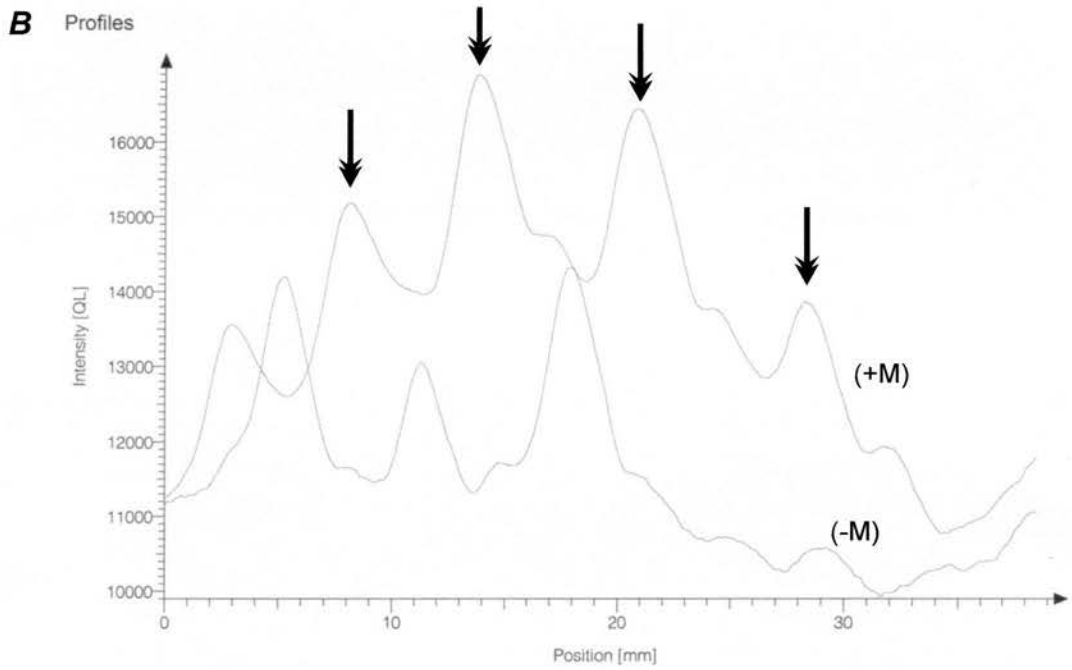


Figure 3.8 *B* Peak determination scan of the Benzonase analysis performed in figure 3.8 *A*. The region scanned from figure 3.8 A is indicated alongside the image by a bracket. Only the region of most pronounced cleavage by Benzonase was chosen for analysis. The peaks were determined using Aida software coupled with a Fugifil PhosphoImager. Lane 2 in the unmethylated fragment was compared to lane 5 in the methylated fragment. Vertical arrowheads indicate regions of enhanced cleavage in the methylated fragment as judged by Benzonase Analysis. (+M), the CpG methylated DNA fragment. (-M), the unmethylated DNA fragment.

specifically. PCR fragments generated from pBR322 were split in half. These are the same fragments used by Mei *et al.,* (1988), that were generated here for benzonase analysis as described in 3.1.2.2 above. Again, one batch was methylated with *Sss*I methylase and checked for successful methylation with a methylation protection assay (data not shown). The batches were γ–^{32}P-labeled at one end and subjected to a benzonase assay in the absence and presence of 4 M urea. The results of this experiment are shown in figure 3.9. In the digestion pattern, unmethylated DNA is compared to methylated DNA in both the standard Benzonase digestion buffer and the same buffer containing 4 M urea. The difference due to methylation is clearly seen in both methylated samples digested and is highlighted in figure 3.9. This pattern does not seem to be enhanced in the urea containing samples. Both methylated digestion patterns seem to have the same frequency of cutting in standard and urea containing samples. There also appears to be the same frequency and pattern of digestion in the unmethylated sample lanes. In conclusion from this experiment there appears to be no advantage in adding urea to the digestion buffer solution. However, to further confirm this result the experiment was repeated as described below.

Fragments generated from pGEM9Zf(-) were prepared as described in 3.1.2.4 above and are the Pingoud fragments used in 3.1.2.4 above. These correspond to the same fragment as that used by Meiss *et al.,* (1999). Benzonase analysis was carried out on these fragments in the absence and presence of 4 M urea. Figure 3.10 *A* shows the results of this experiment. Again, in each methylated sample, the digestion pattern clearly reveals a methylation dependent change at the same region as that shown in figure 3.8. As this analysis showed a clearer pattern of the methylation enhanced cleavage, the region of most pronounced cleavage corresponding to the d(CGCGCGCG)$_n$ sequence was selected for peak determination analysis. The results of this analysis are shown in figure 3.10 *B.* In fact this image provides a more accurate view of the methylation induced structural changes, especially over the B-forming alternating d(CG)$_4$ sequence, indicated by the *Hha*I digest in marker lane M. In the unmethylated sample this is poorly reacted, but as soon as it is CpG methylated it becomes a highly efficient cutting site. This methylation-induced

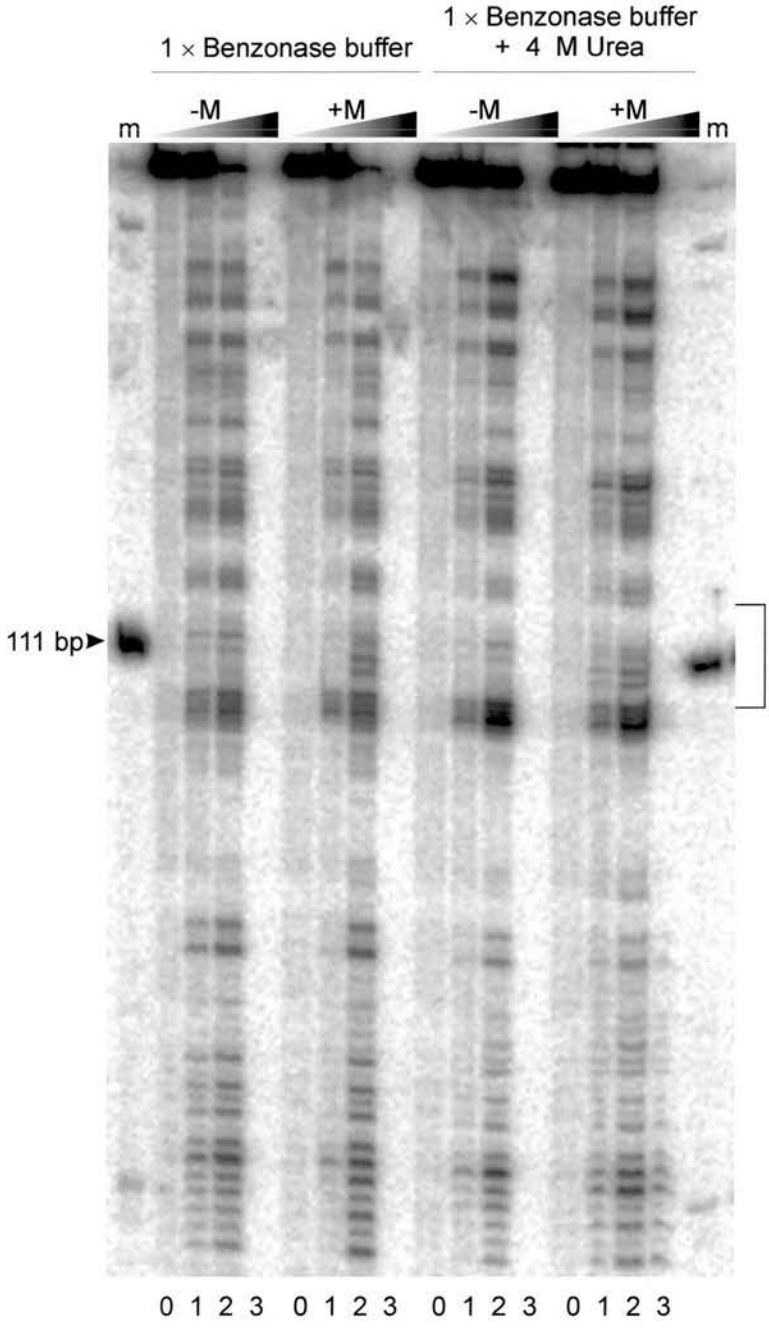


Figure 3.9 Benzonase digest of pBR322 fragments in altered conditions. Benzonase digest of PCR generated fragments from pBR322 in standard Benzonase reaction buffer conditions (1X buffer Y^+, MBI), compared to standard Benzonase reaction buffer containing 4 M Urea. Fragments were *Sss*I methylated (+M) and unmethylated (-M). **m,** *Hha*I digest of unmethylated pBR322 fragment. The bracketed region shows the change in structure due to methylation. This is the same methylation induced change as seen in either (+) or (-) Urea. It is also the same shift in pattern as that seen in figure 3.5.

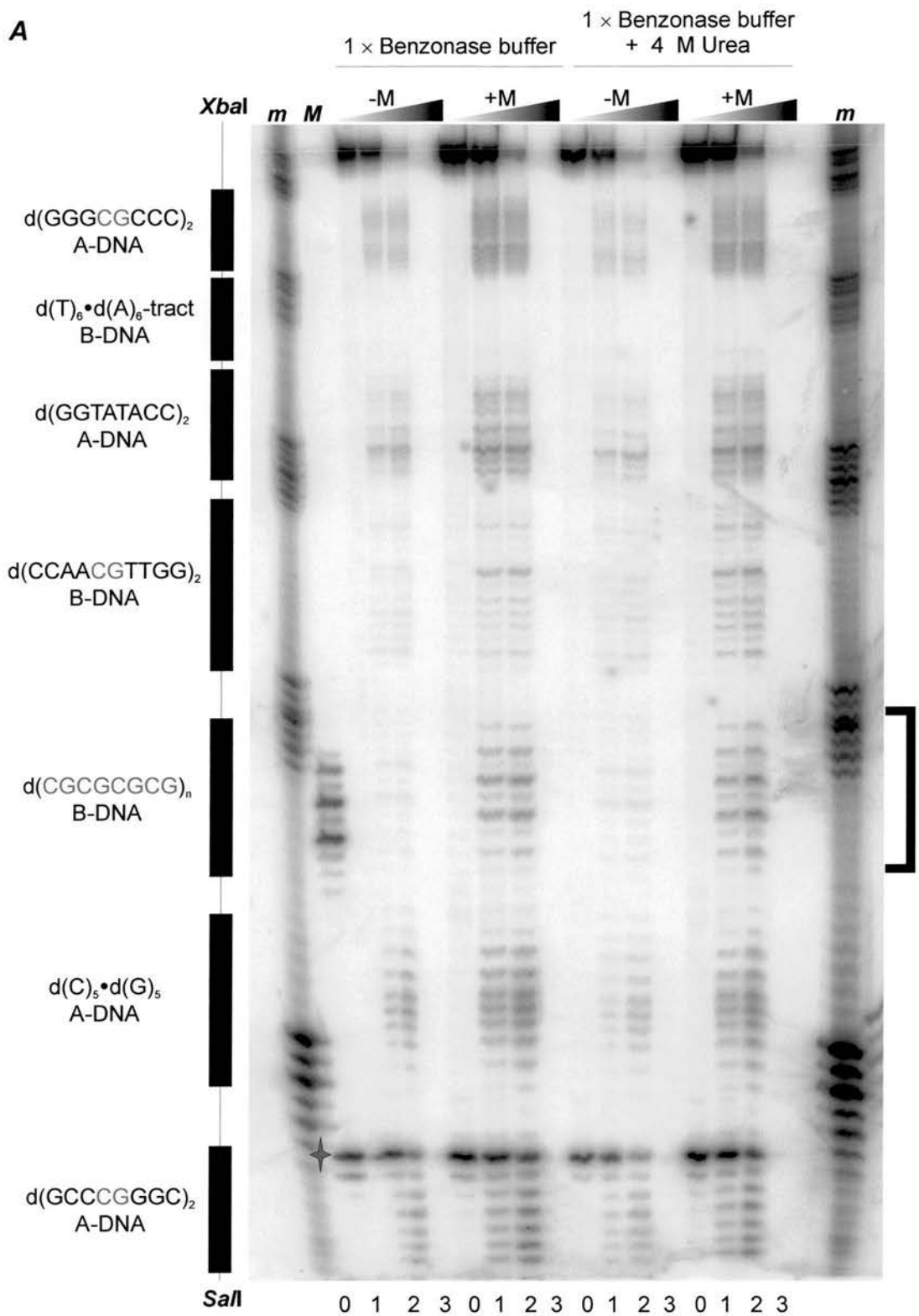


Figure 3.10 *A* Benzonase digest of pGEM9Zf(-) fragments in altered conditions. Benzonase digest of PCR generated fragments from pGEM9Zf(-) in standard benzonase reaction buffer conditions (1 × Buffer Y$^+$, MBI) compared to standard benzonase reaction buffer containing 4 M urea. Fragments were *Sss*I methylated (+M) and unmethylated (-M). ***m***, 20 bp DNA Ruler (Bio-Rad); ***M***, *Hha*I digested unmethylated pGEM9ZF(-) fragments. The star symbol indicates an anomalous band present at this position across the gel. The sequences within the fragment are indicated by bars on the left hand side of the image.

B

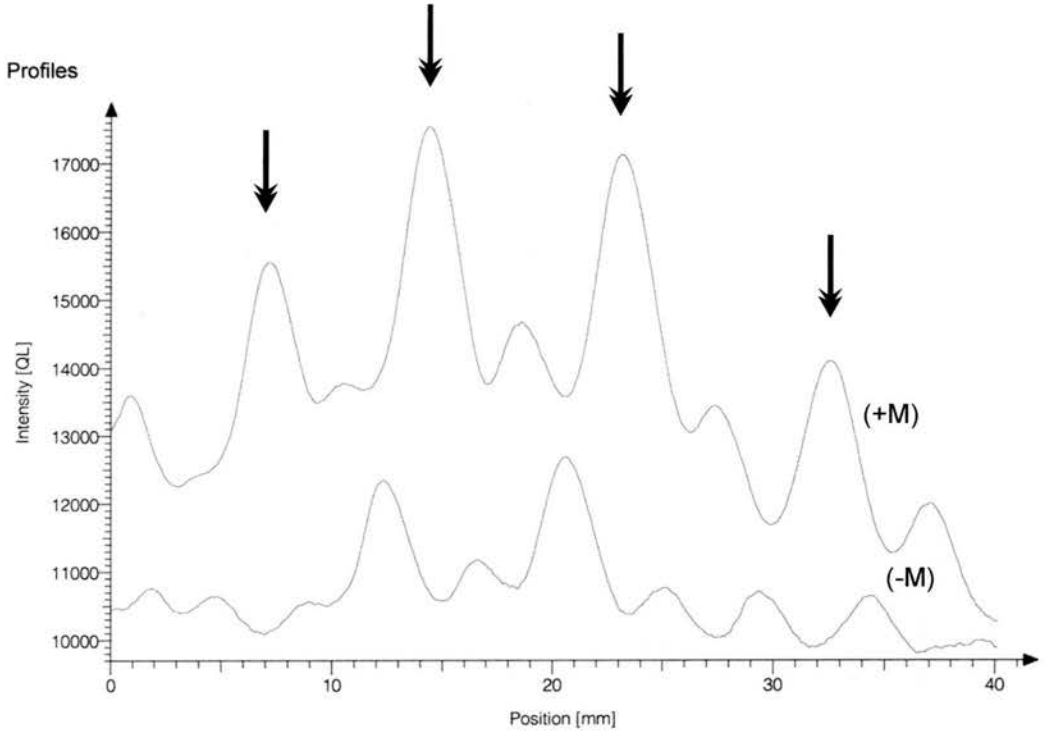


Figure 3.10 *B* Peak determination scans of the Benzonase digestion assay shown in figure 3.10 *A*. The region scanned is highlighted in figure 3.10 with a heavy bracket alongside the gel image. This corresponds to the sequence d(CGCGCGCG). There are four methylatable sites present in the sequence and it appears that when methylated this sequence shows enhanced cleavage with Benzonase. Lane 2 of the methylated fragment was compared to lane 2 of the unmethylated strand for the 4M urea treated lanes. (+M) refers to the methylated fragment. (-M) refers to the unmethylated fragment. Vertical arrows indicate sites of methylation enhanced cleavage as judged by Benzonase analysis.

observation was made here and was not a feature of any of the analyses by Meiss *et al.*, (1995, 1998, 1999). However in this experiment, there does not seem to be any increased specificity due to the presence of 4 M urea. The unmethylated samples also show the same patterns of digestion in both standard and 4 M urea containing reaction buffers. In conclusion, although Meiss *et al.*, (1995) have reported increased specificity for Benzonase in the presence of urea, coupled with the reduction of cleavage at non A-forming sequences, by my analysis I have not seen the same effect. In fact, there appears to be no advantage in adding urea to the reaction buffer, and herein I have eliminated it from the standard protocol to keep the reaction conditions as close to physiological conditions as possible.

3.1.2.6 DNA methylation and globin gene regulation

Having established the effectiveness of Benzonase for the analysis of CpG methylated DNA, the possible function and relevance of methylation induced A-form DNA structures was investigated further by selecting a CpG rich gene known to be affected *in vitro* by DNA methylation. For this, the human α_2-globin gene was chosen and later the chicken β^A-globin gene was analysed.

There is a strong correlation between CpG methylation of both α- and β-globin gene control regions and transcriptional activities in different tissues (Ginder *et al.*, 1998). Generally, CpG methylation of globin gene promoter regions correlates with transcriptional inactivity at developmental stages or in tissues where the genes are not expressed (Singal *et al.*, 1997). Abnormalities in the normal methylation status of globin genes has been linked directly to a variety of disorders: the thalassemias, diabetes and other blood conditions (review: Weatherall, 2001).

The correlation between transcriptional regulation and methylation of the human α–globin locus is a little less clear than for the β-globin cluster. The β-globin genes lack CpG islands and behave as typical tissue-specific genes. In contrast, although the alpha-globin genes are also erythroid-specific, they are 60 % CpG rich and ***do*** contain CpG islands. The alpha-globin genes also lack scaffold-attachment regions, a feature that the β-globin genes possess (review: Higgs *et al.*, 1989).

CpG islands are highly G+C rich regions of non-methylated DNA associated with about 50% of mammalian gene promoters They remain non-

methylated throughout development regardless of the expression of the gene, with the exception of CpG islands in the inactive X-chromosome and those at imprinted genes. All house keeping genes and many tissue-specific genes fall within this group. The other 50% of mammalian genes have promoters with a G+C content and methylation levels indistinguishable from bulk DNA. These are invariably associated with tissue-specific genes (Antequera & Bird, 1993).

It has been long known that in many non-erythroid cell lines, the human α_2-globin gene is not expressed, whether the CpG island is methylated or not. However, recent RT-PCR analysis of the human α_2-globin gene has revealed that in some non-expressing cell lines, where the CpG island is unmethylated, expression is not completely eliminated, but reduced forty-fold. In cell lines where the CpG island, extending into the promoter ***is*** methylated, α_2-globin gene expression is repressed further (Cuadrado *et al.,* 2001). There may be a specific type of chromatin environment in different cell lines that actively represses genes not required for proliferation. Indeed many cell cycle genes are also repressed by methylation in cell lines. Therefore, protein factors alone may reduce expression substantially, but CpG island methylation can completely abolish it. It is probable in this case, that the coordinated activation and repression of the globin gene families is via methyl-CpG mediated transcriptional mechanisms. Indeed, when the CpG islands of these genes are found methylated, in HeLa cell extracts - they can bind to MeCP1 (Boyes & Bird, 1991). The exact organisation and control of these and other mechanisms surrounding globin gene expression is presently not understood.

3.1.2.6.1 Preparation of human α_2-globin gene fragments for Benzonase analysis.

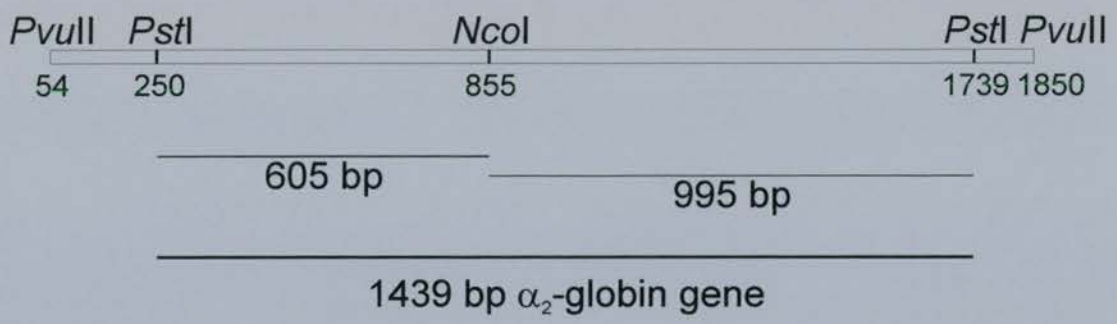
A pUC9 plasmid containing the full-length cloned version of the 1.5 kb human α_2-globin gene inserted into the *Pst*I site (figure 3.11), was kindly provided by Prof. Adrian Bird. Fragments of the gene were prepared by various methods. Initially, plasmids were transformed into *E. coli* JM110 cells and grown to an O.D. of 0.8. The plasmids were recovered by midipreparation and purified. Samples were taken and differentially methylated. Three methylating enzymes were used: (a) *Hha*I methylase (MBI) which methylates the central CG in the sequence GCGC, (b) *Hpa*II

methylase (MBI) which methylates the second cytosine in CCGG and (c) *Sss*I methylase (NEB) with methylates every CpG. To confirm complete methylation of the fragments they were then taken and individually challenged with methylation sensitive and insensitive enzymes. ie: *Hha*I methylated fragments were challenged with *Hha*I restriction enzyme. *Hpa*II methylated fragments are challenged with *Hpa*II and *Sss*I methylated fragments are challenged with a mixture of both *Hha*I and *Hpa*II enzymes. Figure 3.12 shows the results of this analysis and is an example of the methylation protection assay used for all methylated DNA analyses in this thesis. They were then digested with *Pvu*II to extract the full-length 1.5 kb gene. It was then digested with *Nco*I producing two fragments, one of 1004 bp and the other of 811 bp. These enzymes are insensitive to CpG methylation or do not contain methylated CpGs in their recognition sites. These fragments were isolated on a 1.0 % agarose gel. They were gel purified and a sample of each recovered fragment run on a 1.0 % agarose gel to check for purity and concentration (data not shown). The 811 bp methylated fragments were then 5'-end labeled with ^{32}P-γ–ATP using T4 kinase. These were chosen as they contain the transcriptional start and promoter region of the α_2-globin gene. They were then cut at a single *Pst*I site at the start of the gene sequence to produce a fragment of 605 bp that is labeled at the *Nco*I end on one strand only. The sequence of this fragment is shown in figure 3.11. These fragments were separated from the end that was removed and then subjected to Benzonase digestion to reveal any regions of methylation-induced structural changes.

3.1.2.6.2 Benzonase analysis of the human α_2-globin gene.

Benzonase digestion of an unmethylated 605 bp α_2-globin gene is compared to an *Sss*I methylated version of the same fragment. Figure 3.11 shows the sequence of the α_2–globin gene used with the methylatable CpGs highlighted. Figure 3.13 shows the results of this assay. The marker lane provides a guide of *Hha*I sites in the DNA sequence. As seen from figure 3.13, there are a number of structural changes throughout the sequence. These are mainly located towards the end of the gel at positions that coincide with the promoter region of the gene sequence. This experiment was repeated several times (data not shown). Each time the results showed the same alterations over this sequence. A peak determination scan for three

A



B

```
CTGCAGGAAG CGAGGCTGGA GAGCAGGAGG GGCTCTGCGC AGAAATTCTT TTGAGTTCCT  60
ATGGGCCAGG GCGTCCGGGT GCGCGCATTC CTCTCCGCCC CAGGATTGGG CGAAGCCCTC 120
CGGCTCGCAC TCGCTCGCCC GTGTGTTCCC CGATCCCGCT GGAGTCGATG CGCGTCCAGC 180
GCGTGCCAGG CCGGGGCGGG GGTGCGGGCT GACTTTCTCC CTCGCTAGGG ACGCTCCGGC 240
GCCCGAAAGG AAAGGGTGGC GCTGCGCTCC CGGGTGCACG AGCCGACAGC GCCCGACCCC 300
AACGGGCCGG CCCCGCCAGC GCCGCTACCG CCCTGCCCGG GCGAGCGGGA TGGGCGGGAG 360
TGGAGTGGCG GGTGGAGGGT GGAGACGTCC TGGCCCCCGC CCCGCGTGCA CCCCCAGGGG 420
AGGCCGAGCC CGCCGCCCGG CCCCGCGCAG GCCCCGCCCG GGACTCCCCT GCGGTCCAGG 480
CCGCGCCCCG GGCTCCGCGC CAGCCAATGA GCGCCGCCCG GCCGGGCGTG CCCCCGCGCC 540
CCAAGCATAA ACCCTGGCGC GCTCGCGGCC CGGCACTCTT CTGGTCCCCA CAGACTCAGA 600
GAGAACCCAC CATGG 615
```

CG = CpG dinucleotide
= CCGG sequences, *Hpa*II restriction sites
= GCGC sequences, *Hha*I restriction sites
= overlaps

Figure 3.11 The human α_2-globin gene sequence. In ***A*** above the complete human alpha globin gene sequence inserted into the pUC9 plasmid vector is outlined. ***B*** The DNA sequence of the 615 bp *Pst*I to *Nco*I gene fragment (inclusive of the enzyme recognition sites), used for Benzonase analysis. The restriction enzyme sites for *Pst*I and *Nco*I are underlined in the text at the 5' and 3' ends of the sequence respectively. The TATA box is also underlined in the text. Other details are explained in the figure. Fragments were labelled at the *Nco*I end and cleaved at the *Pst*I end in the Benzonase digests shown. GCGC sequences are *Hha*I sites which have a strong potential to form A-DNA when CpG methylated in previous Benzonase digests. CCGG sites are *Hpa*II sites as indicated in the diagram.

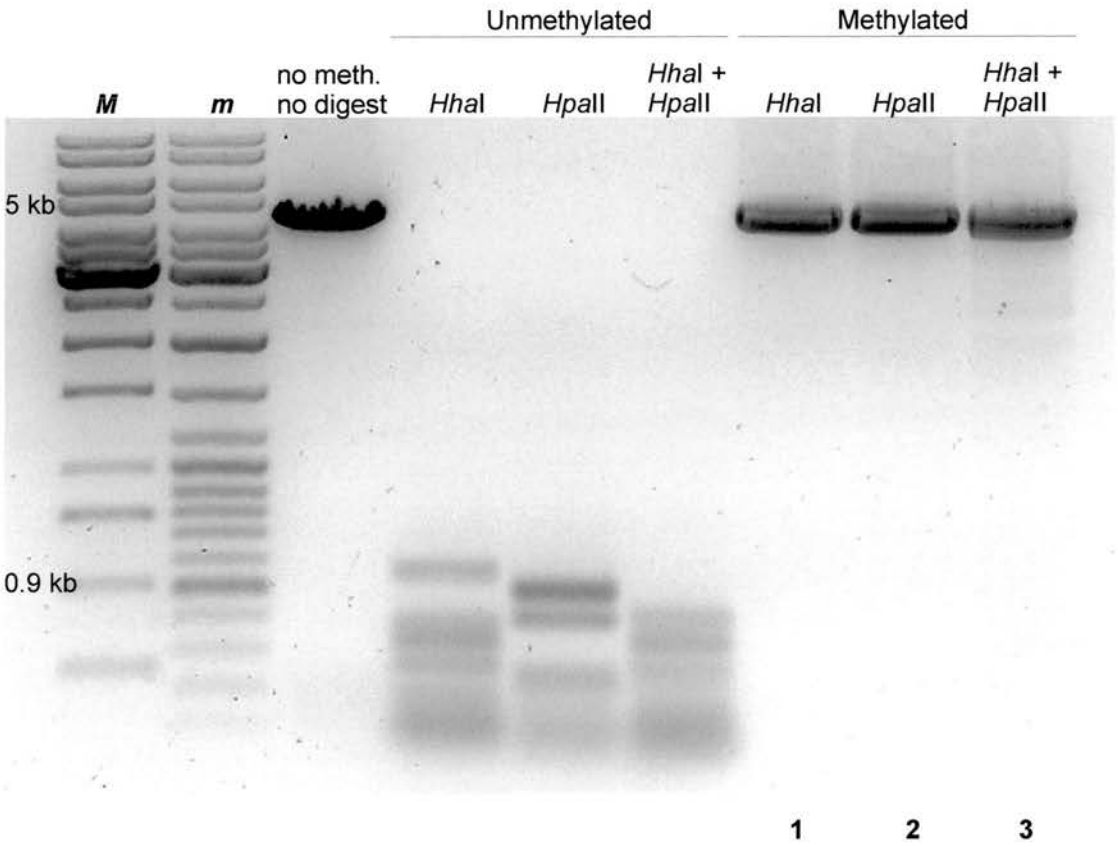


Figure 3.12 Methylation protection assay of pUC9 plasmid containing the human α_2-globin gene. To assess that fragments are fully methylated prior to Benzonase , DNaseI and footprinting analysis- they are subjected to a methylation protection assay. *M,* 1 kb DNA ladder mix (MBI-fermentas); *m,* pBR322 digested with BsuRI (MBI-fermentas). no meth. no digest, Unmethylated plasmid and undigested control. Unmethylated plasmid is digested by *Hha*I, *Hpa*II and a combination of *Hha*I and *Hpa*II. These are control lanes to demonstrate both the activity of the enzymes and the digestion pattern of unmethylated plasmid DNA. Methylated plasmid is also subjected to restiction digests with *Hha*I, *Hpa*II and a mixture of *Hha*I and *Hpa*II. 1, the plasmid has been *Hha*I methylated; 2, the plasmid has been *Hpa*II methylated; 3, the plasmid has been *Sss*I methylated.

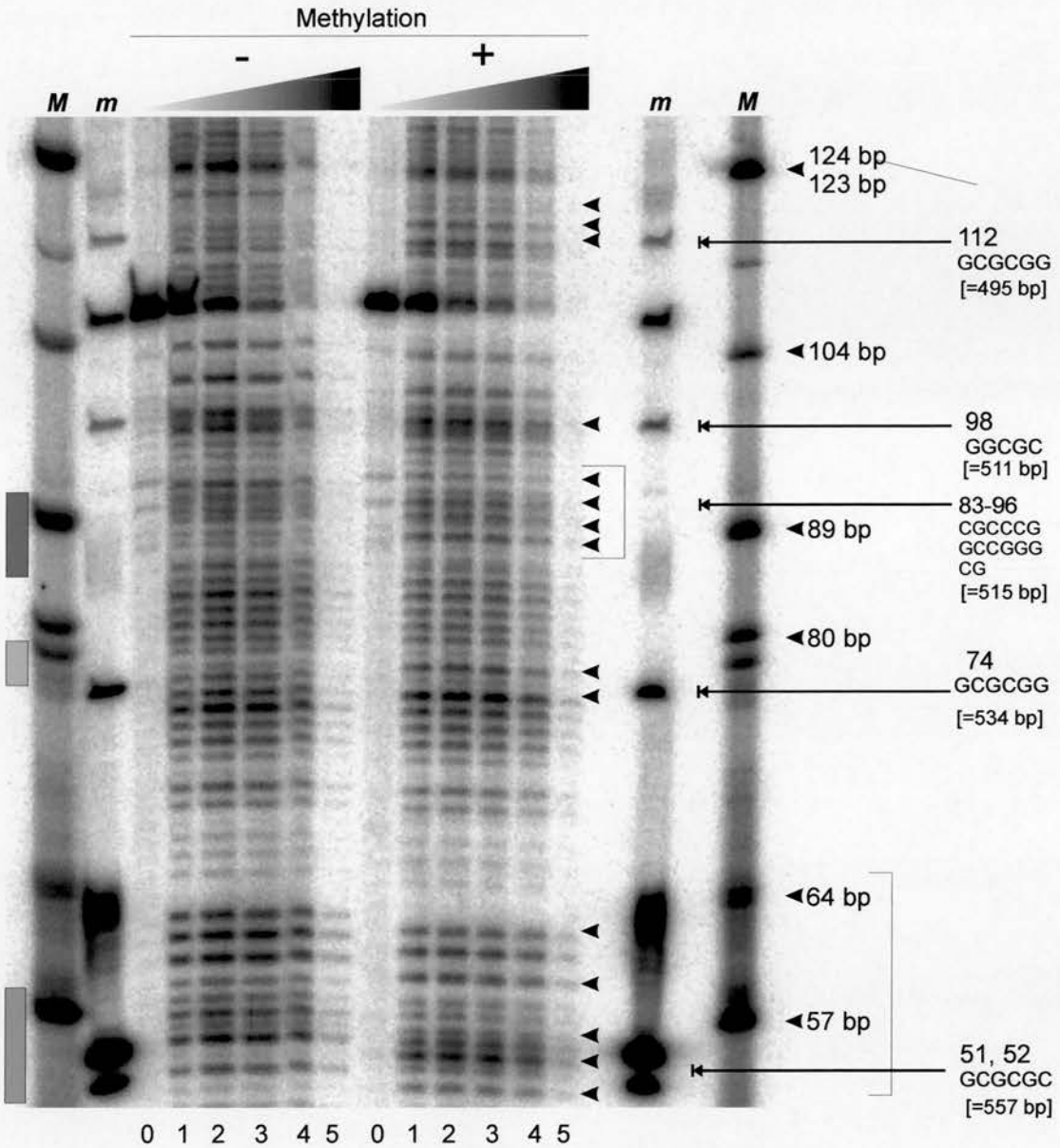


Figure 3.13 *A* Benzonase analysis of the human α_2-globin gene. Unmethylated (-) and *Sss*I methylated (+) α_2-globin DNA fragments were benzonase digested. Small arrowheads in the lane ***M*** indicate marker sizes; small arrowheads on the gel indicate regions of methylation induced DNA structural change. ***M***, pBR322 digested with *Bsu*RI; ***m***, α_2-globin gene fragment digested with *Hha*I. The bracketed region shows a region of a large amount of methylation induced structural change. Long arrows indicate some of the *Hha*I sites where Benzonase shows enhanced cleavage at CpG methylated sites. The region marked between 83 and 96 bp contains many CpGs, some in *Hpa*II sites and some in partial *Hha*I sites. It is unclear whether methylation induced structural changes at this position are due to the *Hpa*II sites or because there are several methylatable CpGs in this region. Coloured bars indicate regions scanned. The values in base pairs (bp) is given in the right hand margin of the individual fragment sizes. In square brackets, [], underneath each value is the corresponding position in base pairs of the sequence, from the 5' end as shown in figure 3.11, page 109.

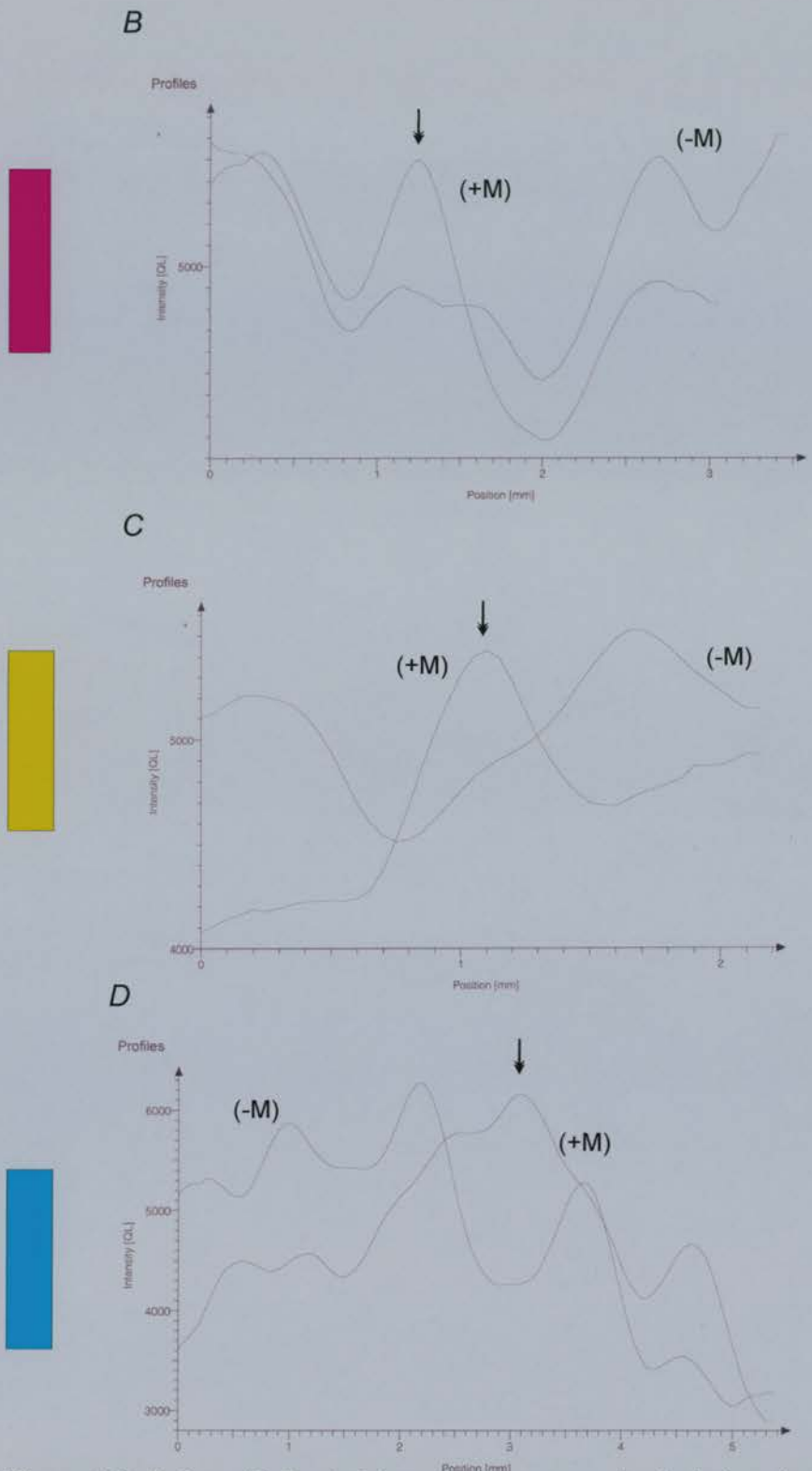


Figure 3.13 *B, C and D.* Peak determination scans of selected regions of the Benzonase assay shown in figure 3.14 *A.* Selected regions were scaned using Aida software. Each scanned region is colour coded with reference to 3.14 A. The intensity values are arbitrary values determined by the range in intensity range available in the data set. In ***B,*** lane 3 (-M) is compared to lane 2 (+M). In ***C & D,*** lanes 3 are compared. (-M), unmethylated. (+M) methylated. Vertical arrows indicate methylation induced peaks in the sequence as detected by Benzonase.

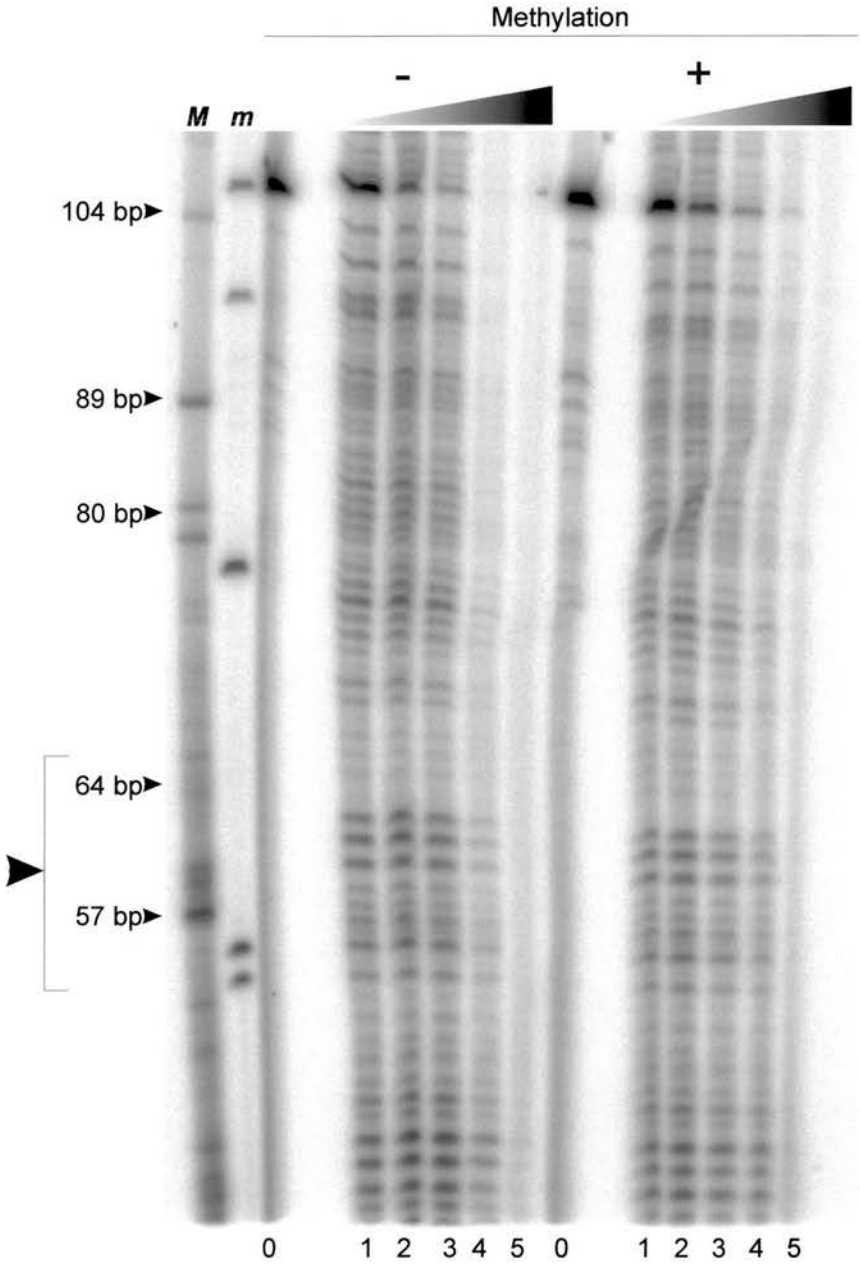


Figure 3.14 Benzonase analysis of the human α_2-globin gene. Unmethylated (-) and *Hpa*II methylated (+) α_2-globin DNA fragments were benzonase digested. Small arrowheads indicate marker sizes; large arrowhead indicates regions where no methylation induced DNA structural changes occur compared to Figure 3.13. ***M***, pBR322 digested with *Bsu*RI; ***m***, α_2-globin gene fragment digested with *Hha*I. This figure illustrates that the methylation induced changes seen in figure 3.13 are likely due to methylation at *Hha*I sites and not at *Hpa*II sites.

selected regions is shown in figure 3.13 *B, C & D.* Over small regions it is clearly seen that methylation is having a major impact on DNA structure.

To confirm that these alterations were due to *Hha*I methylase target site changes, *Hpa*II methylated fragments of the same sequence were subjected to Benzonase analysis. The results of this experiment are shown in figure 3.14. Although the banding patterns are not as defined as some other gels, it is clear that there are no obvious structural transitions due to methylation of *Hpa*II sites as detected by this method. Therefore most Benzonase reactive structural transitions due to CpG methylation may be due to reactivity at *Hha*I sites. The *Hha*I methylase methylates the sequence d(GCGC), an alternating GC stretch of nucleotides, the structure of which has been shown to potentiate the formation of A-DNA when methylated (table 1.2). The *Hpa*II methylase recognizes the sequence CCGG, which is non-alternating and not shown to potentiate the formation of A-DNA. The *Hpa*II methylated fragment is an indicator that these methylation induced structural changes are mainly due to methylation at *Hha*I sites.

Clearly there are many structural alterations and clearly there are many methylatable sites in the DNA as seen from the DNA sequence. This experiment has been repeated and the same structural alterations are observed (data not shown). Clearly there are no observable changes in the *Hpa*II sequence, indicating that the methylation induced structural changes, as detected by Benzonase, are low or non-reactive, unless within other alternated DNA tracts.

3.1.3 Summary

This chapter introduces the Benzonase assay for analysis of CpG methylated DNA structure. This method demonstrates the existence of potential A-form DNA regions in CpG methylated DNA *in vitro*, a feature that may have functional consequences in terms of gene regulation in the cell.

An initial hypothesis was based on the fact the CpG methylation can induce A-form DNA (Vargason *et al.*, 2000, 2001; Tippin *et al.*, 1997a & 1997b). Experiments using plasmid DNA and digesting with Benzonase, an enzyme with a preference for A-form DNA cleavage, revealed that the digestion rates of

unmethylated and *Sss*I methylated DNA fragments were similar. Therefore, CpG methylation alone is not enough to alter the digestion dynamics of the whole sequence (figure 3.2) but may be having some effect at the level of the base pair.

The Benzonase assay as developed by Meiss *et al.*, (1995) was applied to CpG methylated DNA. Here, radiolabeled fragments of DNA are digested and run on a denaturing PAGE/urea gel. In figure 3.5 an unmethylated fragment of the pBR322 plasmid was compared to a methylated fragment of the same sequence of the same fragment. In an accompanying experiment (figure 3.6), DNaseI analysis was performed on the same fragments. Peak determination scans of both of these gels was performed and the results shown in figures 3.5 C and 3.6 C, respectively. Several conclusions can be drawn from these analyses. (1) The fragments used here were also used in a study to detect A-DNA with a chiral probe specifically designed to detect A-DNA. The same sequences show enhanced activity with Benzonase confirming that Benzonase is as reported, a useful probe for A-DNA detection. (2) A CpG methylated sequence at 111 bp shows high reactivity with Benzonase when compared to the unmethylated sequence. This corresponds to an alternating sequence containing two methylated CpGs, d(CCGCGC) that has been crystallised as A-DNA (Mooers *et al.*, 1995; Tippin *et al.*, 1997). (3) Other methylated CpGs do not show heightened Benzonase reactivity even though some of these are surrounded by other cytosine and guanine nucleotides. However, none of these other sequences show alternation of more than three base pairs and none of them contain more than one meCpG. Some of them may correspond to E-DNA, an intermediate in the transition to A-DNA that can be formed from a single methylation site in the sequence d(GGCGm^5CC), (Vargason *et al.*, 2000). This may indicate that Benzonase requires a more pronounced change to A-DNA than the intermediate structure allows to demonstrate hyper-reactivity. (4) DNaseI analysis confirms that every CpG is methylated and that methylation is capable of altering DNA structure at each of these sites. It also establishes that Benzonase is capable of ignoring these structural changes in favour of more pronounced changes that suit its active site.

As another test, the gene fragment originally used by Meiss *et al.*, (1999) in studies to determine the nuclease activity of Benzonase was used to confirm their observations that Benzonase had a preference for A-DNA. This proved to be the case

(figure 3.8 *A, B*). However, when the sequence was methylated, the pattern changed and showed altered cleavage patterns, now preferring to cut at a methylated CpG string of nucleotides, a feature consistent with the above observations.

Attempts to refine the Benzonase assay were made in accordance with Meiss *et al.*, (1995). urea was used in the reaction buffer in an effort to enhance cleavage at potential A-form regions, while reducing cleavage at non A-form sites. Figures 3.9 and 3.10 *A &B,* illustrate that there is no difference in the cleavage patterns observed in both unmethylated and methylated fragments in standard and urea containing buffers. The structural changes due to methylation are still clearly seen in both cases, but the addition of urea has not enhanced reactivity at specific sites. urea was not used in any future Benzonase digests, as it was not deemed useful and detracted from the near physiological conditions of the assay. However, it did also demonstrate that the structural observations made were still observable in such altered conditions, indicating a stable change in the DNA, one that is intrinsic to the sequence and not an alteration due to environmental influences.

However, bacterial sequences do not predominantly use CpG methylation for gene regulation in the same way as eukaryotic cells, nor do they contain a higher order chromatin environment of regulatory proteins. To assess a gene that may use these structural alterations a region of the human α_2-globin gene was analysed. This gene is highly CG rich and is repressed by DNA methylation *in vitro* and *in vivo* (Boyes & Bird, 1992; Cuadrado *et al.,* 2001). Benzonase analysis of this gene shows an altered pattern of digestion at many *Hha*I sites corresponding to the sequence d(GCGC), predominantly at the promoter region of the genes where many regulatory proteins involved in transcription, translation and the regulation of gene expression may bind (figure 3.13). However, Benzonase could not detect any methylation induced changes of the same sequence when methylated with *Hpa*II methylase, which methylates the sequence d(CCGG). This is shown in figure 3.14.

To conclude it appears from these initial analyses, that the degree of A-DNA formation as judged by the reactivity of Benzonase, is determined by the degree of alternation of base sequence, ie: alternating strings of purines and pyrimidines are preferred substrates. CpG methylation promotes the formation of these alternative structures, but may require at least two adjacent methylated CpGs to do so. Thus we

may have A-DNA type structures embedded within stretches of standard B-DNA, Z-DNA or indeed other stable intermediate structures. The existence and interchanging of these structures has currently no determined role in the eukaryotic cell.

The α_2-globin gene provides a good starting point for finding a distinct biological role for methylation induced A-type DNA structures. The next layer of information above the level of DNA sequence is the compaction of DNA by nucleosomes. However, the α_2-globin sequence has not been characterized in a chromatin context. In the lab of Dr. Jim Allan, a technique for mapping the exact location of nucleosomes on a DNA sequence to the level of the base pair, was developed (Davey *et al.*, 1995). Mapping studies using the technique of monomer extension have determined the position of nucleosomes on the chicken β^A-globin gene promoter region (Davey *et al.*, 1997). In some cases, nucleosome positioning could be altered by CpG methylation and was hypothesized to be due to an altered DNA structure. Because of the wealth of knowledge about the sequence and nucleosomal organisation of this gene, I have taken it as a model for the determination of a potential role for A-DNA in the regulation of gene expression. The β^A-globin gene is the subject of analysis in Chapter 4: Results (II).

Chapter 4: Results (II)

4.1 Analysis of the chicken β^A-globin gene

4.1.1. Introduction

In eukaryotic nuclei, the bulk of DNA is packaged into nucleosomes. Nucleosomes consist of histone octamers containing histones H2A, H2B, H3 and H4. DNA is folded and compacted by nucleosomes into a higher order chromatin structure, the 30nm fibre (figure 1.2), which is organised into regulatory domains (review: van Holde, 1994). The 30nm fibre is the substrate that all eukaryotic DNA binding proteins are faced with. The manner in which nucleosomes temporally and spatially organise themselves to construct these domains is not fully understood. It is known that the DNA sequence has some influence (reviews: Simpson, 1991; Thoma, 1992; Wolffe, 1994).

As discussed in chapter 1, the mechanisms by which CpG methylation could influence gene expression at a particular site in DNA are conceptually threefold. (1) CpG methylation could occlude protein binding sites directly either by altering a recognition sequence and/or by changing the structure of a protein binding site. (2) CpG methylation can affect nucleosome positioning, with knock-on effects for gene regulation, and (3) it could affect gene expression by interacting with MeCPs and subsequent chromatin remodeling complexes. All of these effects are plausible and examples exist of each. We are currently learning a great deal about the third possibility, but little about the first and second.

In relation to nucleosome positioning, it has been shown that core histone octamers reconstituted onto the promoter of the chicken β^A-globin gene, adopt characteristic translational positions with respect to the underlying DNA (Yenidunya *et al.*, 1994; Davey *et al.*, 1995). Spatial organisation of nucleosomes may be of prime importance in the developmental organization of this gene. In terms of DNA structure, Davey *et al.*, have investigated the influence of CpG methylation on nucleosome positioning on the chicken β^A-globin gene promoter. Methylation of a CpG triplet was found to directly influence the position of an otherwise high affinity

nucleosome binding site. This CpG triplet is located 1.5 helical turns away from the dyad axis, orientated towards the histone core (Davey *et al.*, 1997). This work established that CpG methylation can influence nucleosome positioning and potentially gene expression at the level of DNA structure (figure 4.1). Here, a very limited number of nucleotides can influence directly the translational positioning of nucleosomes. DNaseI analysis of the β^A-globin gene in unmethylated and CpG methylated states, showed enhanced cleavage at many methylated CpGs, including the methylated triplet, indicating methylation dependant structural changes (Davey *et al.*, 1997). Reasons for this were speculated on. Methylation may have influenced the bending potential and subsequently secondary or tertiary structure of the DNA, thus turning an otherwise high affinity nucleosome binding site, into a low affinity binding site. It has also been demonstrated that a CpG triplet can adopt a Z-DNA conformation under negative superhelical stress (Runkel & Nordheim, 1986). This transition is further enhanced by CpG methylation. Based on this knowledge, Dr. C. Davey analysed the CpG triplet in the β^A–globin gene for the formation of Z-DNA (pers. comm.). However, Z-DNA was found only to form on supercoiled DNA and not when the DNA fragment was linearised (unpublished results). It was reasoned that Z-DNA could not be responsible for the movement of nucleosomes away from the methylated CpG triplet *in vitro,* as nucleosome movement can occur on the linear methylated β^A-globin CpG triplet, when Z-DNA cannot form there (pers. comm.). Hence, from discussions with Dr. C. Davey it was reasoned that the β^A-globin gene promoter region may be a good candidate for A-DNA analysis, as the specific region in the promoter region that alters nucleosome positioning consists of an alternating sequence containing a CpG triplet. The alternating sequence begins with a Guanine, and is of the type 5'-GGCGCGCGC-3', a sequence shown by crystallization studies to form A-DNA. Mooers *et al.*, (1995) demonstrated that both Gm^5CGCGC and Gm^5CGm^5CGC sequences form A-DNA. Tippin *et al.*, (1997) also demonstrated that the sequence $Gm^5CGm^5CGCGCGC$ crystallizes as A-DNA. In collaboration with Dr. C. Davey, I decided to analyse it using the Benzonase assay that I had optimized for A-DNA analysis on CpG methylated DNA.

Concurrently, Dr. C. Davey decided to further analyse the structure of the β^A-globin promoter region in the wild-type form and in a variety of point mutants (m1,

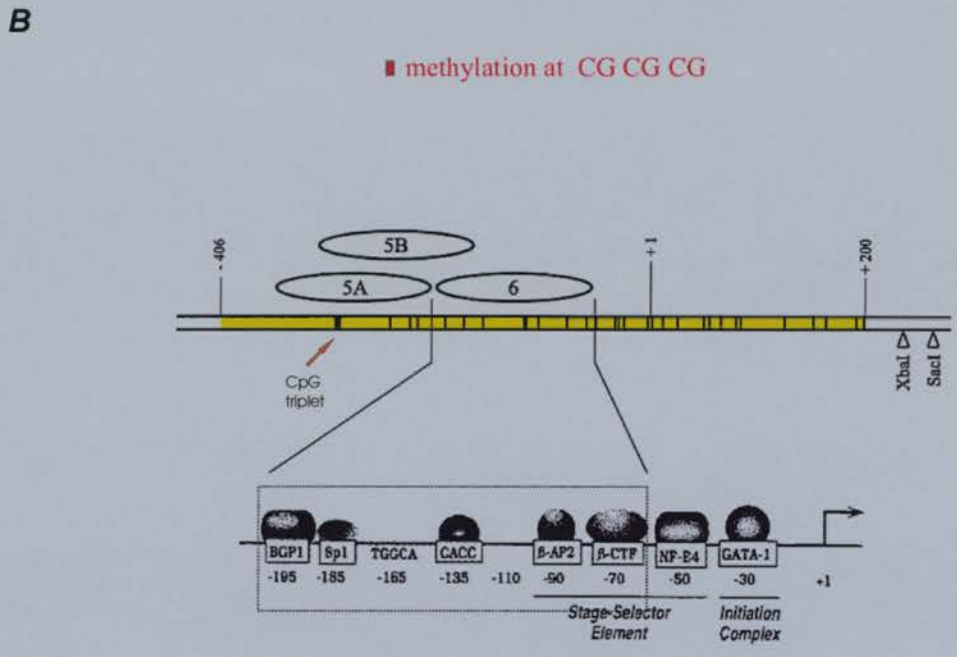


Figure 4.1 CpG methylation alters nucleosome positioning. ***A*** The far left illustration represents the location of nucleosomes centrally and at the end of a piece of DNA. The central image is a polyacrylamide gel of nucleosomes reconstituted onto DNA showing that CpG methylation (lane 5) alters the positioning of nucleosomes 5A and 5B (compare lane 3) The far right image is lane peak analysis of lanes 3 & 5 from the PAGE image. Subtraction of these peaks (d) shows that the positioning of nucleosome 5A has been moved by CpG methylation. ***B*** This figure illustrates nucleosome positioning over the β^A-globin gene promoter region, the positioning of nucleosomes 5A and 5B on this fragment and below, different factor binding sites across the promoter region (Adapted from Davey *et al.*, 1995, 1997). It is the region covering this promoter that I have also used in my analyses.

	Sequence
Wild-Type (LE)	5'...GGCGCGCGCTGTGCT...3'
Mutant No. 1 (m1)	GCCGCG
Mutant No. 2 (m2)	CGGCCG
Mutant No. 3 (m3)	CGCGGC
Mutant No. 23 (m23)	CGCCCG
Mutant No. 25 (m25)	CGGGCG
Mutant No. 8a (m8a)	CGGCCGCTGTGCG

Table 4.1: Site-directed mutagenesis in the β^A-globin CpG triplet. The wild-type sequence and mutated versions of the β^A-globin triplet sequence. The mutants were prepared by site-directed mutagenesis by Dr. C. Davey. In red, CG nucleotides. In black, wild-type nucleotides. Underlined are the point mutation bases as compared to the wild-type sequence.

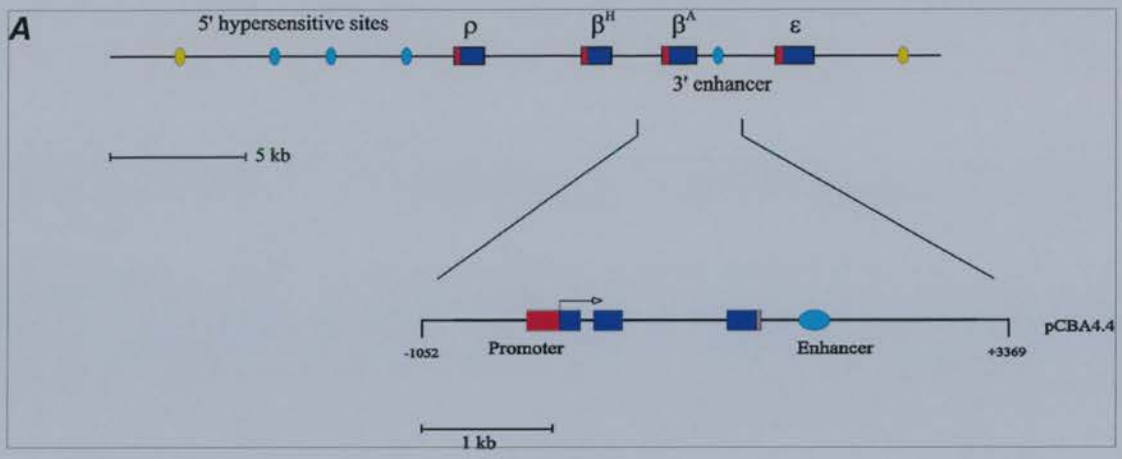
m2, m3, m23, m25 and m8a) that he prepared (table 4.1). On these, both DNaseI and nucleosome positioning studies were performed, to investigate both DNA secondary structure and nucleosome positioning respectively. I have taken the same range of mutants to analyse their behavior when subjected to A-DNA analysis. The results of Dr. C. Davey's experiments are discussed in relation to my own data in support of an argument that A-form DNA has a role to play in the positioning of nucleosomes on DNA and potentially in the regulation of gene expression.

4.1.2 Results

4.1.2.1 β^A-globin gene Fragments

To investigate the potential formation of A-DNA in the chicken β^A-globin gene, plasmids containing cloned inserts of the gene were kindly provided by Dr. C. Davey. These included wild-type β^A-globin (referred to as LE) and a selection of mutant variants (referred to as m1, m2, m3, m23, m25 and m8a) containing point mutations in the CpG triplet (table 4.1). These were each transformed into competent *E.coli* JM110 cells and grown to an O.D. of 0.8. The cells were harvested and plasmid midipreps prepared. In all cases, plasmids were digested with *Cla*I and *Hinf* I (NEB) restriction enzymes. The resultant fragments were resolved on a 5% polyacrylamide gel (data not shown). A band of 219 bp was isolated comprising the promoter region of the β^A-globin gene, and containing the CpG triplet analysed previously (Davey *et al.*, 1995, 1997). Figure 4.2 illustrates the β^A-globin DNA fragment isolated and the DNA sequence thereof. The most significant restriction enzyme cutting sites within the 219 bp sequence are marked. This band was purified by dialysis in an electrophoresis tank according to the method of Dr. C. Davey. The concentration of the isolated fragments were estimated by analysis on a 2% agarose gel in reference to marker DNA of known concentration (data not shown). The remainder of the DNA samples were then dephosophorylated at their 5' ends with Bacterial Alkaline Phosphatase (BAP) or Calf Intestinal alkaline Phosphatase (CIAP), (MBI-fermentas).

Many preparations of these fragments were prepared for the initial studies and repeat experiments to verify the results. In some cases, the method of preparation of fragments was altered. For instance, some of the mutant 219 bp



C

1 bp

5'-CGATAAGCTTGATCTGGTGTGCTGGGAGGAAGGACCCAACAGACC
CAAGCTGTGGTCTCCTGCCTCACAGCAATGCAGAGTGCTGTGGTTTGG
AACTGTGTGAGGGGCACCCAGCCTGGCGCGCGCTGTGCTCACAGCAC
TGGGGTGAGCACAGGGTGCCATGCCCACACCGTGCATGGGGATGTAG
GCGCACTCCGGTATAGAGCTGCAGAGCTGGG-3'

219 bp

Figure 4.2 The chicken β globin locus. ***A*** The organisation of the β^A-globin gene cluster with the organisation of the β^A-globin gene highlighted (Image drawn to scale, supplied by Dr. C. Davey). ***B*** The region of the β^A-globin gene used in this study. Numbering refers to the base pair positions relative to the transcriptional start site.(Image not drawn to scale). ***C*** The nucleotide sequence of the 219 bp fragment used in this study. All CG dinucleotides are highlighted in red. The CpG triplet is roughly located at the centre of the sequence and is underlined.

fragments were prepared by a quicker method as follows. These plasmids were digested similarly, but isolated on a 1.8 % agarose gel. The bands corresponding to the 219 bp fragments were cut out, placed in eppendorf tubes containing 1 ml of gel extraction buffer and placed on a rotating wheel at 4°C overnight, to allow the DNA to diffuse out of the gel pieces. The DNA was purified by standard phenol-chloroform extraction and ethanol precipitation. The precipitated DNA was washed with 80 % ethanol to remove excessive salts.

In all cases the concentration of the DNA was again estimated by running 5 % of the final volume on a 1.8 % agarose gel. For quantification of the isolated DNA fragments a phosphorImager (Fugifilm) was used to scan the gel and Aida™ software used to estimate the concentration as accurately as possible. Spectrophotometry at A_{260} was also used to corroborate these results (data not shown). These fragments were dephosphorylated with Shrimp alkaline Phosphatase (SAP) instead of BAP or CIAP as it is active at 37°C and can be completely destroyed by incubation at 65°C for 15 minutes.

The dephosphorylated 219 bp fragments (LE, m1, m2, m3, m23, m25 and m8a) were then subjected to CpG methylation. Samples of each wild type and point mutant were taken and split in half. One half of each sample was methylated using *Sss*I methylase (NEB), which has the ability to methylate DNA at every CpG site on both DNA strands. All DNA samples subjected to *Sss*I methylation were then tested for their complete methylation at CpG sites, by a methylation protection assay. On every occasion that DNA was methylated, a methylation protection assay was carried out to confirm the methylation status. In the assay, CpG methylated DNA is challenged by use of *Hha*I and *Hpa*II restriction enzymes (MBI). Both have recognition sites containing CpG dinucleotides, but both will not cut if their recognition sites are CpG methylated. Unmethylated DNA was used as a control to test the integrity of the *Hha*I and *Hpa*II restriction enzymes. As an additional control, the restriction enzyme *Msp*I was also used. It recognizes the same sequence as *Hpa*II, but unlike *Hpa*II it can cut methylated DNA (Data not shown). Once the DNA fragments were satisfactorily methylated they were again quantified.

4.1.2.2 Benzonase Analysis of β^A-globin Wild-Type (LE) fragments (+ Strand).

Samples of unmethylated β^A-globin wild-type (LE) fragments were compared to samples of methylated β^A-globin wild-type (LE) fragments in Benzonase assays. These *Cla*I-*Hinf*I fragments contain the wild-type version of the CpG triplet analysed by Davey *et al.*, (1997). This is of the sequence 5'...GGCGCGCGCTGTGCT...3'. Highlighted in yellow is the CpG triplet. In blue, the alternating nucleotides, $(GC)_4$. The fragments were ^{32}P-γ-ATP labeled at their 5'-ends with T4 kinase. The double-stranded fragments were restriction digested with *Pst*I to remove one of the labeled ends. The fragments were then digested with Benzonase over an increasing concentration gradient according to the standard devised protocol. The resultant fragments were purified by phenol-chloroform extraction, ethanol precipitated and and run on an 8% denaturing polyacrylamide gel. This is illustrated in Figure 4.3 *A* and 4.3 *B*).

Figure 4.3 *C* shows a peak-determination scan of the highlighted regions in figure 4.3 *B*. These peak determinations were performed using Aida™ software. It is clearly visible from this graph that the banding pattern is incredibly similar over the length of the selected regions, but differs significantly over the CpG triplet region. Significantly, three peaks are visible over the CpG triplet, representing the three methylated CpG sites.

A pattern emerges showing a striking overall similarity in digestion patterns between unmethylated and methylated β^A-globin fragments (figure 4.3 *A*). Apart from the overall similarity, increased cutting is seen at hotspots in the DNA sequence that possess an ability to form A-DNA (figure 4.3 *B*). At the CpG triplet d(GCGCGCGC) located at 124 bp in the DNA sequence, a marked change is observed across lanes 0 to 6 in the 5'-CpG methylated fragment only, clearly visible as marked in the peak determination scan (fig. 4.3 *C*). Here, it is evident that Benzonase now recognizes a new hotspot in the sequence, indicating the formation of A-form DNA. At this site the sequence is methylated at two overlapping *Hha*I *methylase* d(GCGC) sites. As seen from Chapter 1, table 1.2, a sequence that begins with a purine start and contains alternating GC sequences is more likely to form A-DNA. In addition to this, three sequences containing 5'-methylated CpG's have been crystallised as A-form DNA. These are d(Gm^{5}CGCGC), d(Gm^{5}CGm^{5}CGC) and

d(Gm^{5}CGm^{5}CGCGCGC). The table also shows that the unmethylated sequence of d(GCGCGCGCGC) forms Z-DNA. It is the only type of alternating DNA with a purine start that crystallizes in Z-form. For sequences of the type d(CGCGCGCG), structural studies show the Z-conformation in both solution and crystalline states, but only in the unmethylated form. All that is required to promote this transition is the addition of four methyl-groups. This startling transition from the left-handed Z-form to the right-handed B-form and A-form structures is one of the most structurally dynamic changes in biology. Provided here may be some evidence for a functional biological role for these transitions.

The CpG triplet in the unmethylated β^A-globin gene is $d(GC)_4$. The closest crystallised match is $(GC)_5$. It is likely that the unmethylated $d(GC)_4$ also forms Z-DNA, however as mentioned previously, this may depend on the degree of supercoiling the sequence is subjected to. When methylated, the sequence may conform to any one of the three methylated purine start oligomers that have been crystallised in A-form (Chapter 1, Table 1.1). There must therefore be an equilibrium established between the formation of A, B, and Z-DNA to enable these transitions to be possible. There is evidence to suggest that this is so from thermodynamic studies (Foloppe *et al.,* 1999), and also evidence to suggest that such transitions in DNA structure may be important in setting up higher order chromatin structure (Ohyama, 2001).

Also evident from figure. 4.3 *B*, is the fact that the methylation not only causes structural changes over the methylated CpG triplet, it also has a small spreading effect. The pattern of cutting is altered at adjacent nucleotides to the methylated CpG triplet. This is not so surprising as the formation of A-DNA is likely to bring torsional strain and twisting to the adjacent nucleotides that do not form A-DNA. This experiment was repeated at least three times more. Each time, the same results and observations were made as above.

DNaseI digestion analysis for the wild-type and mutant fragments is shown in figure 4.5 (data from Dr. C. Davey). Nucleosome positioning data for the wild-type and mutant fragments is shown in figure 4.6 *B*. Both the DNaseI and nucleosome positioning data were performed on the same fragments used in my analyses. The

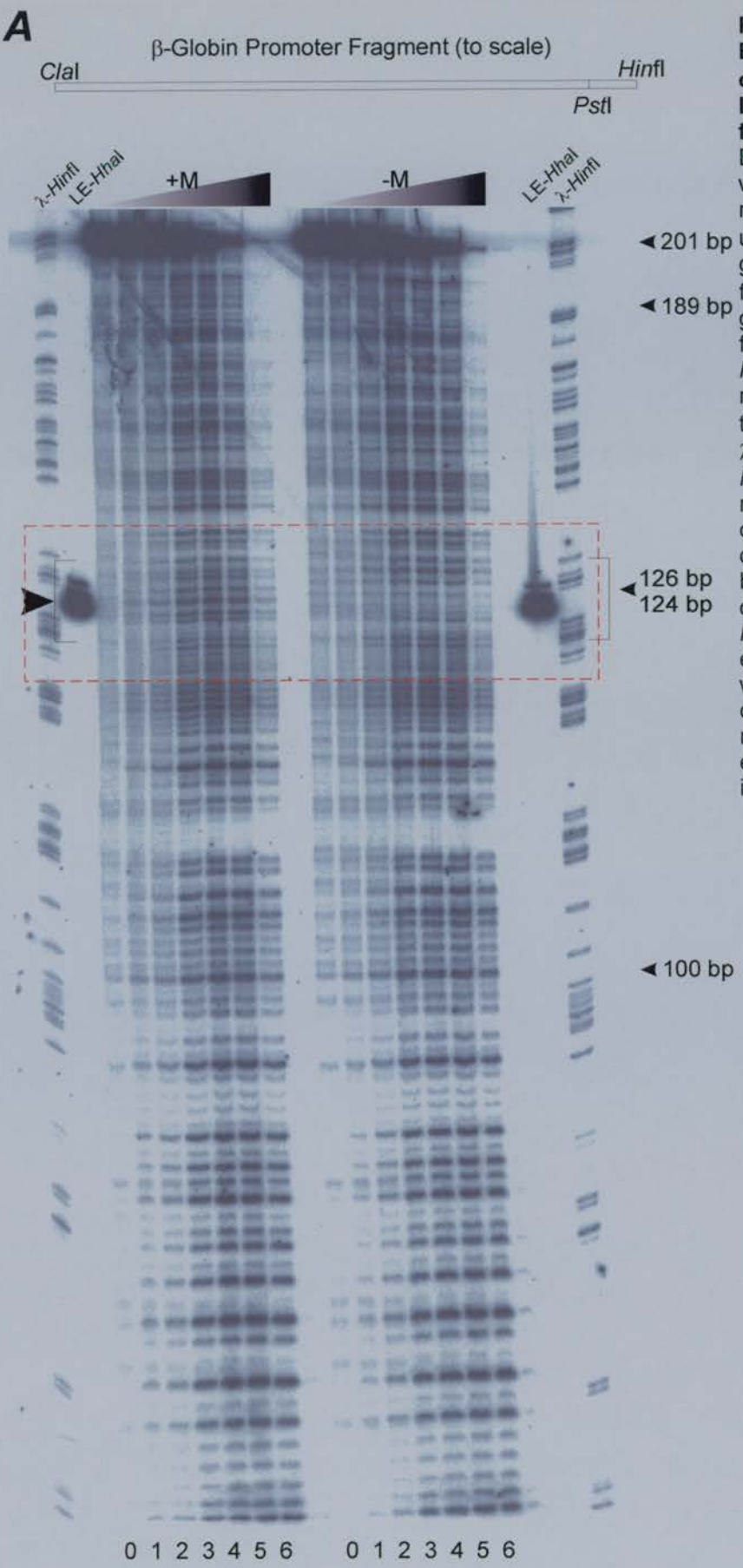


Figure 4.3 A Benzonase analysis of β^A-globin wild-type LE fragments (in forward strand). ***a*** Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet. λ-*Hin*fI, λ DNA digested with *Hin*fI. Arrows represent the positions of methylatable CpG dinucleotides. The bracketed region covers a methylated *Hha*I site that shows enhanced cleavage with Benzonase. The dashed box covers a region that is expanded for viewing in Fig. 4.3 *B*.

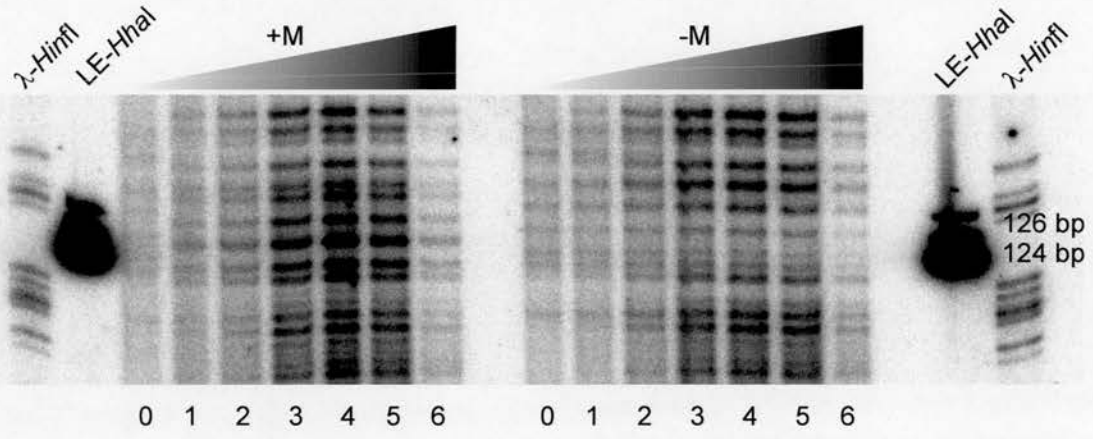


Figure 4.3 *B* Magnified image of highlighted section from Figure 4.3 *A*. This image corresponds to the red-boxed area in figure 4.3 *A*. It clearly shows that at the region of the CpG triplet as marked by the *Hha*I digested fragment, there are hypersensitive sites in the Benzonase digest in the +M lanes. These hypersensitive sites do not occur in the -M lanes. The hypersensitivity is therefore due to a methylation dependant change in structure at the CpG triplet, an effect that is not seen at other methyl CpGs in the fragment.

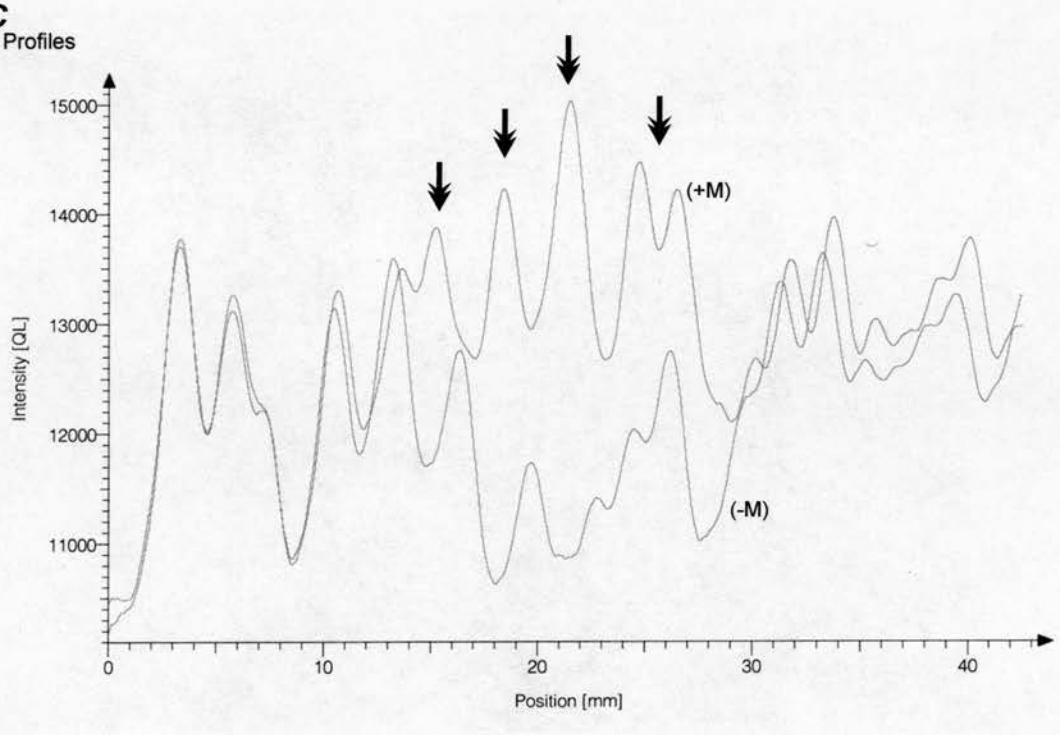


Figure 4.3 *C* Peak determination scan using Aida™software. Aida™software was specifically designed by Fugifilm for the purposes of PhosphoImager analysis. Having scanned PAGE/Urea gels using a Fugifilm PhosphoImager, the data was analysed using lanes of equal radiolabel intensity to establish any differences between unmethylated and methylated samples of the globin fragment in *A* and *B* above. The intensity is in arbitrary units from the Aida software and based on the intensity level range for the entire data set. The position refers to the position in the gel from the starting reference point of the analysis. The region used is bordered by red box. The upper and lower vertical limits were used. (+M), methylated fragment. (-M), unmethylated fragment.

only difference is that labeling was performed on the reverse strand. However, Dr. Davey observes the same effect on nucleosome positioning on the forward strand and the same effects with methylation induced reactivity in the DNaseI experiments. I also show here that the methylation induced structural change as detected by Benzonase is also present on the reverse strand. The technique used for nucleosome positioning is called monomer extension. A brief summary of these analyses is that DNaseI analysis detects alterations in methylated DNA structure at the CpG triplet. Monomer extension analysis determines nucleosome-positioning sites on DNA. This change in structure at the CpG triplet disrupts the positioning of nucleosome 5A from an otherwise strong positioning site. The nature of the CpG induced structural change is likely to be due to formation of an A-type DNA structure at this region, as detected by Benzonase analysis.

4.1.2.3 Benzonase Analysis of β^A-globin Wild-Type (LE) fragments (- Strand).

As seen from Figure 1.5, A-DNA is a symmetrical structural transition. That is, when A-DNA forms on one strand of a duplex, it must also exist on the other strand. This sets it apart from other structural transitions, such as secondary structure changes that may affect one strand and not the other. Examples include bends, kinks and cruciform structures. To prove that the structural alterations observed are a symmetrical event, the same fragments of the β^A-globin gene were prepared, dephosphorylated and 5'-labeled with ^{32}P-γ-ATP at both ends. This time however, the opposite end of the DNA was cut. Instead of using *Pst*I, the restriction enzyme *Sau*3A1 was used to remove one of the ends. Both *Sau*3AI and *Pst*I enzymes were specially chosen because: (a) their recognition sites lie close to either end of the 219 bp fragment, therefore eliminating only a minor part of the β^A-globin fragment, and (b) they are not enzymatically blocked by CpG methylated DNA in or near their recognition sites. The labeled reverse strand contains the sequence 5'-AGCACAGCGCGCGCC-3'. Again, in blue are highlighted the alternating nucleotides, with the yellow box containing the CpG triplet. Unmethylated and methylated fragments were benzonase digested as before and samples run on an 8% denaturing polyacrylamide gel. This is illustrated in figure 4.4 *A*. As the sample was

digested with *Sau*3AI at the opposite end, the resultant fragment is a different length to the forward labeled fragment. Therefore in the Benzonase digestion assay the appearance of the altered structure appears at a different location on the gel. The marker lane shows the position of the CpG triplet on the forward strand, and a lower band corresponding to the CpG triplet on the reverse strand, where the altered pattern in the methylated lane is highlighted.

It is evident from comparing the patterns of methylated and unmethylated DNA in figure 4.4 *A* and *B*, that here again a distinct structural change is picked up by Benzonase. This is clearly seen when comparing lane 3 in the unmethylated strand to lane 3 in the methylated strand. The change is again localized over the methylated CpG triplet. It again shows that there is a spreading effect into the adjacent nucleotides, which also exhibit altered cutting patterns. This experiment was repeated twice more, each time a similar shift in banding patterns was observed, each time over the methylated CpG triplet only.

Corroborating what is visible from the gel, the lane peak determinations were again analysed using Aida™ software. As can be seen from figure 4.4 *C,* there are distinct differences between unmethylated and methylated LE fragments. These occur at a different location on the gel, as in this case the reverse strand is being analysed. However, these alterations are again coincident with the CpG triplet sequence when measured from the *Hinf* I labeled end.

As the effects seen are due to analysis of methyl-groups on one strand of the double-helix at a time, the fact that the events are symmetrical make it even more likely that methylation is promoting a large structural change, evident on both strands and most likely, involving tertiary structure. As Benzonase has a preference for A-form DNA it is reasonable to suggest that this is indeed what it is recognizing in CpG methylated DNA.

Also of interest, is the fact that not every methyl-CpG in the 219 bp fragment becomes a 'hot-spot' for Benzonase digestion. Singly methylated CpG sites do not seem to form a Benzonase hotspot, but at least two adjacent methylatable CpG sites are required for Benzonase to preferably cut at that site *in vitro*. The following conclusions can be drawn; (a) CpG methylation alters DNA structure in a sequence dependent manner, (b) the event is symmetrical, occurring on both positive and

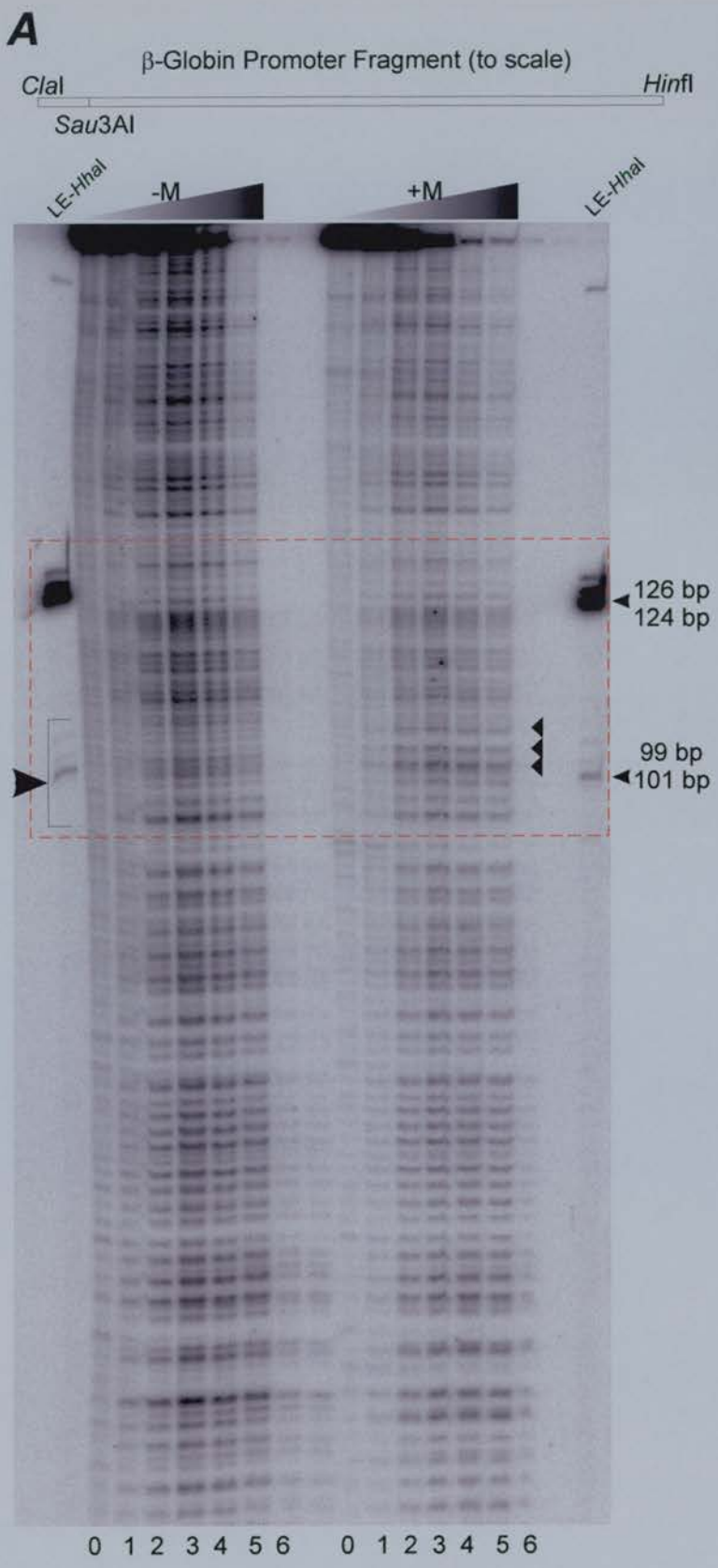


Figure 4.4 *A* Benzonase analysis of β-globin wild-type LE fragments (in reverse strand). *a* Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. These fragments were labelled on the reverse strand as opposed to the forward strand, which was removed by restriction digest using *Sau*3AI. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet on the forward strand at position 126 and 124 bp, and at position 99 bp and 101 bp for the reverse strand. Small arrowheads indicate regions of methylation induced changes in DNA structure as detected by Benzonase.

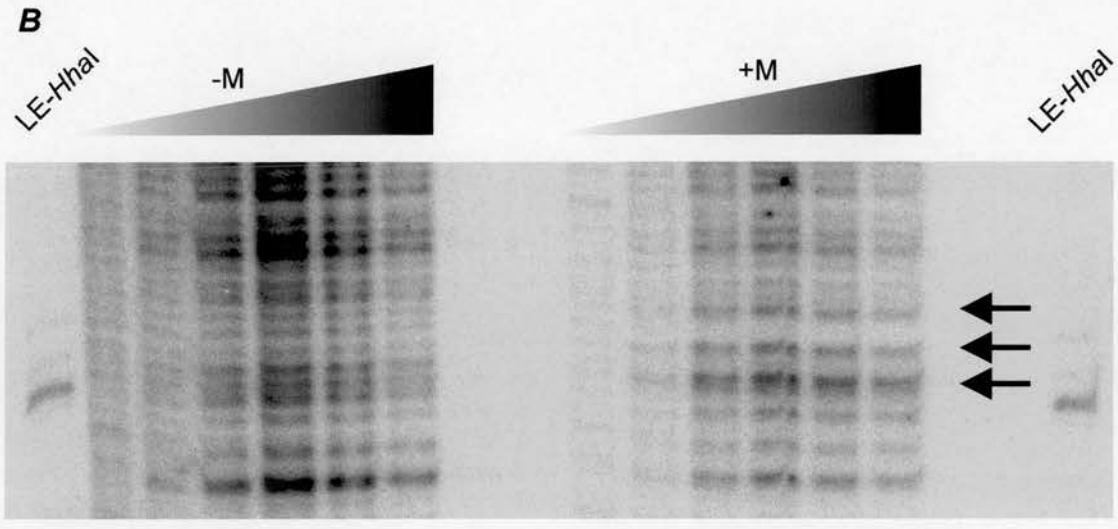


Figure 4.4 B. Enlarged region from Figure 4.6 *A* (red dashed box). Magnified view of the Benzonase digest on the β^A-globin promoter region of the labelled reverse strand. This illustrates that methylation induced cleavage as detected by Benzonase, occurs of both strands of the DNA helix. The arrows indicate the regions of methylation induced hypersensitivity. LE-*Hha*I is the labelled reverse strand fragment digested with *Hha*I. The reverse strand has *Hha*I sites at positions 99 and 101 bp. This coincides with the CpG triplet and with the methylation induced structural change.

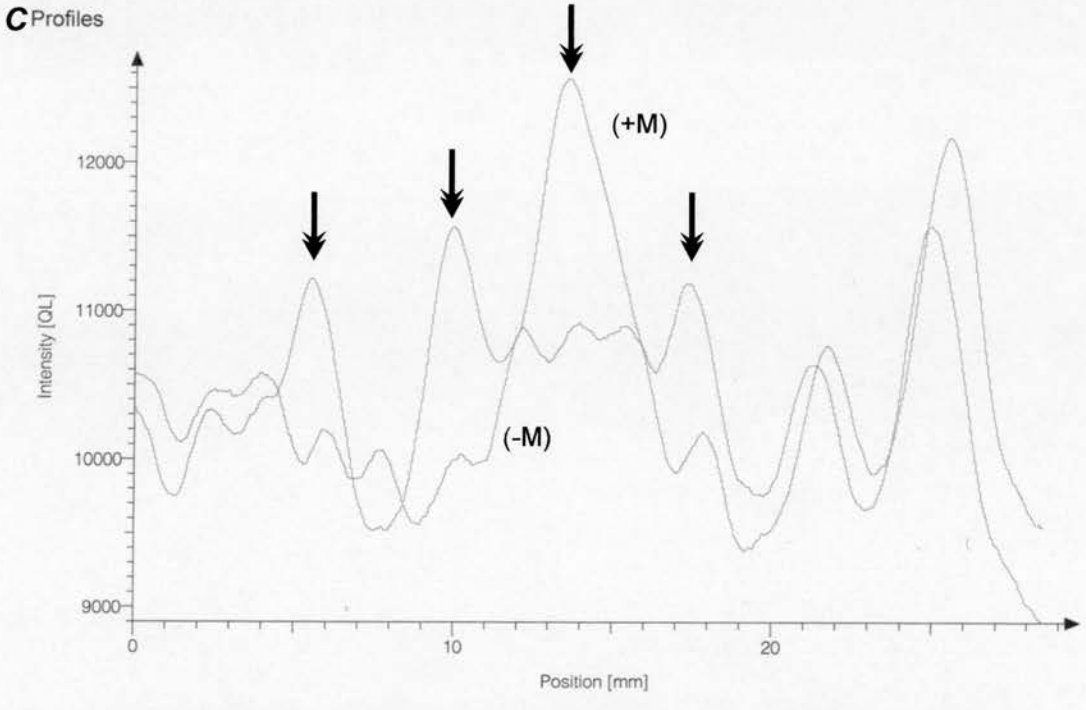


Figure 4.4 *C* Peak determination scan of the Benzonase digestion analysis shown in figure 4.4 *A*. Regions scanned for the β-globin reverse strand analysis are indicated in 4.4 *A* by a bracket. Lane 5 for the unmethylated fragment is compared to lane 4 for the methylated fragment. Regions where CpG methylation has induced a distinct structural change, as detected by Benzonase, are indicated by an arrow. The intensity values along the Y axis are arbitrary and assigned by the Aida software over the full scale of intensity values available in the data set. The X-axis refers to position within the gel over the region scanned. (+M), methylated fragment. (-M), unmethylated fragment.

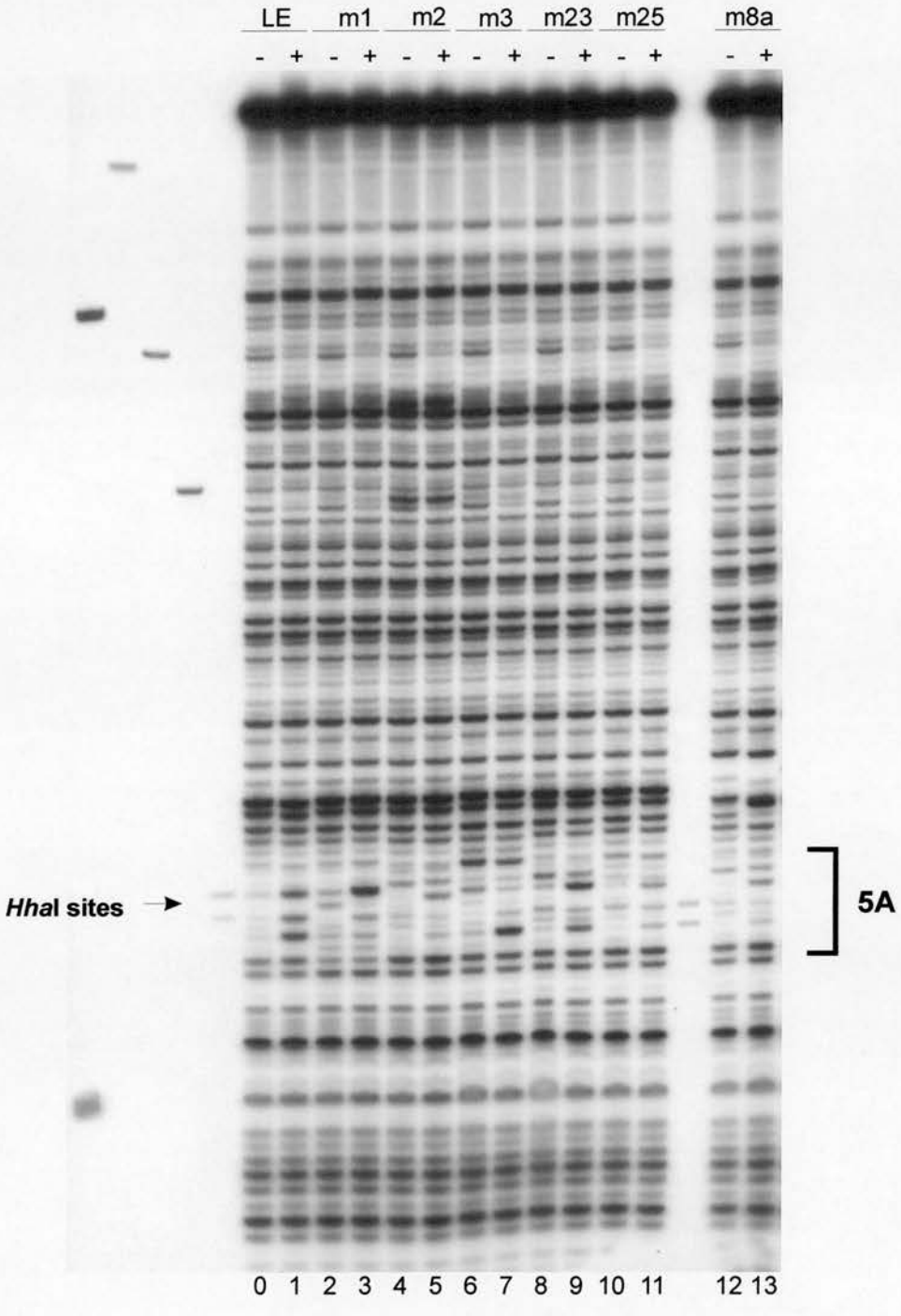


Figure 4.5 *A* DNaseI analysis of β^A-globin promoter fragments. DNaseI analysis was performed to identify DNA secondary structural alterations on unmethylated (-) and methylated (+) wild-type β^A-globin (LE) and mutants (m) 1, 2, 3, 23, 25 and 8a. The region of nucleosome 5A positioning is bracketed. This is also the region of DNaseI detected structural alterations. The *Hha*I sites representing the CpG triplet are marked with an arrow. (This data was provided by Dr.C. Davey and annotated and interpreted here).

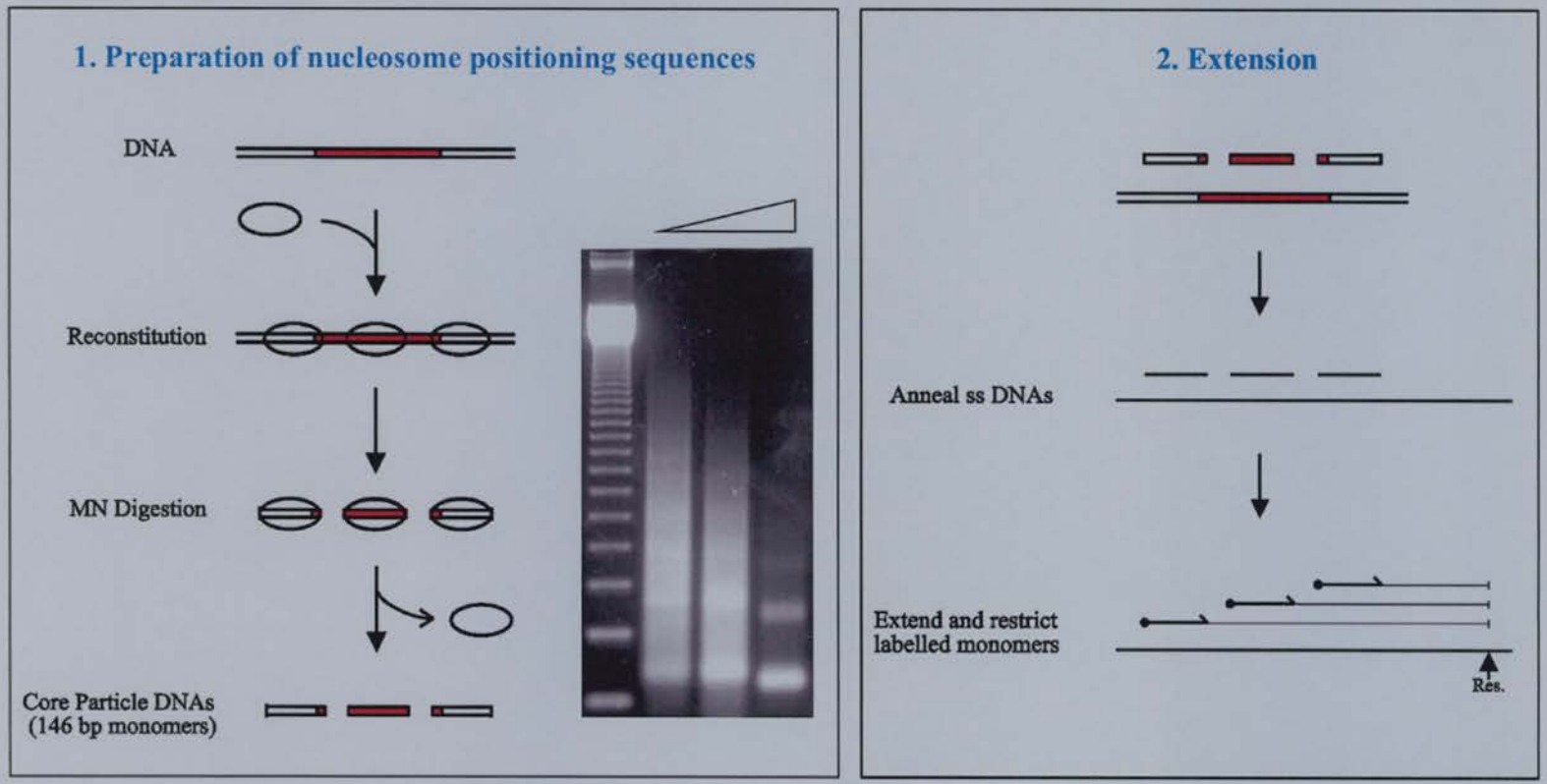


Figure 4.6 ***A*** **Monomer Extension: A method to identify nucleosome positioning sites in DNA.** This method provides long-range, high-resolution, quantitative data on the positioning sites of nucleosomes on DNA. **1**. The production of monomers, sites on DNA where nucleosomes are present. Nucleosomes are reconstituted onto DNA *in vitro.* The intermediate DNAis digested with micrococcal nuclease (MN), avoiding nucleosome bound DNA. The nucleosomal DNA is purified away from bound nucleosomes. The gel shows a timecourse for MN digestion of chromatin. **2.** The isolated fragments are rendered single-stranded, 5'-end labelled with ^{32}P-γ-ATP and annealed to a single stranded template of the original undigested sequence. The annealed fragments act as primers for PCR. All extension products are cut at a known restriction enzyme site. The sites of nucleosome occupancy in the original sequence can now be determined by the length of the fragment. The degree of occupancy is determined by the amount of radiolabel detected for a particular fragment. The extended fragments can be separated according to size on a denaturing PAGE/Urea gel. Exposure to X-ray film will identify the location and signal strength of each band. (Adapted fron Davey *et al.*, 1995, 1997).

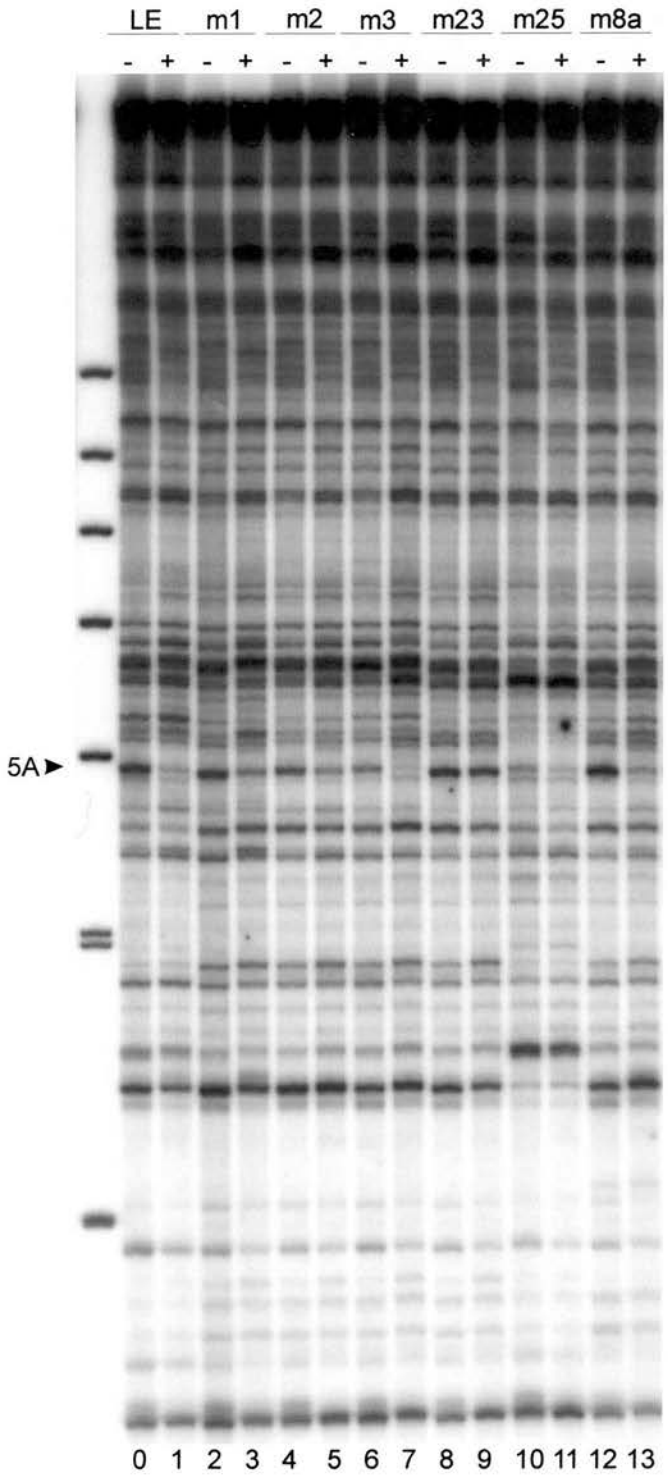


Figure 4.6 *B* Monomer extension analysis of β^A-globin promoter fragments. Monomer extension analysis was performed to identify nucleosme positioning sites on unmethylated (-) and methylated (+) wild-type β^A-globin (LE) and mutants (m) 1, 2, 3, 23, 25 and 8a. The position of nucleosome 5A is clearly seen as indicated by the arrowhead. This data was supplied by Dr. C. Davey.

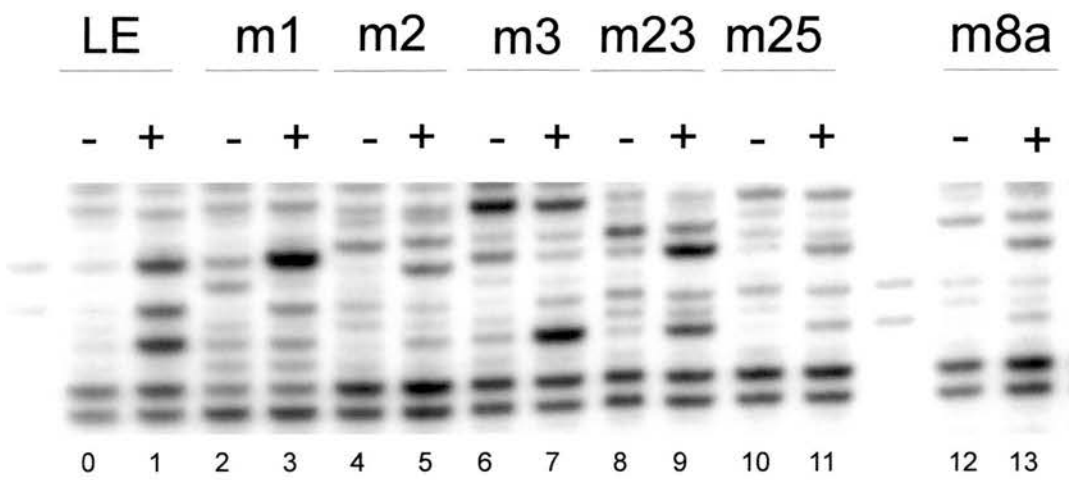


Figure 4.5 *B* Magnified image of the DNaseI analysis shown in figure 4.5 *A*. This image shows a magnified view of the region of the CpG triplet that shows enhanced cleavage. (-) and (+) are unmethylated and methylated lanes respectively. See imge 4.5 *A* for further information. (Unpublished data from 4.5 *A* supplied by Dr. C. Davey).

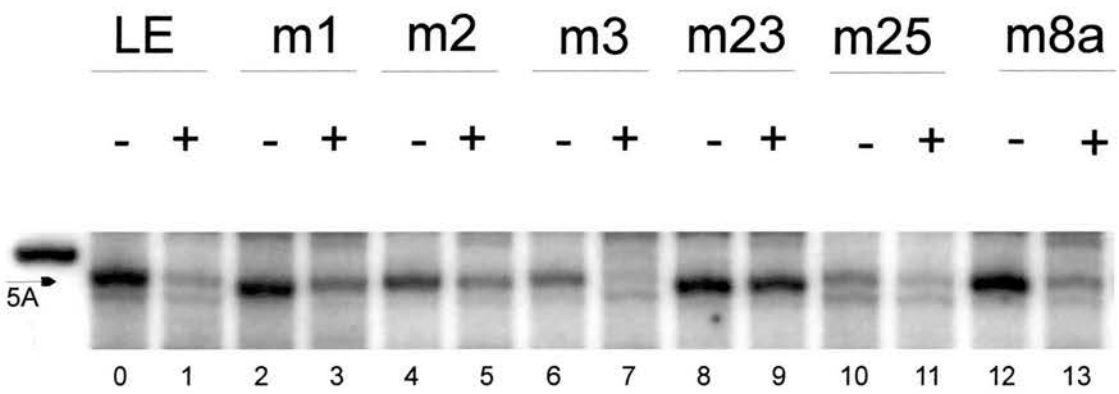


Figure 4.6 C Magnified view of the Nucleosome Positioning analysis shown in figure 4.6 B. This image shows a magnified view of the monomer extension analysis performed as described in figure 4.6 A, the results of which are shown in figure 4.6 B. The bands represent the positioning of nucleosome 5A at the CpG triplet region in the LE fragments and mutant series. See the text for more details. These images are supplied together for ease of analysis. (Unpublished data from figure 4.6 *B* supplied by Dr. C. Davey).

negative DNA strands, consistent with a tertiary structural change (c) in both DNaseI and Benzonase digestion, increased cutting was observed over the me-CpG triplet and (d) in the Benzonase digest experiments, a spreading effect is observed from the me-CpG triplet into the adjacent nucleotides, apparent from the altered cleavage pattern.

4.1.2.4 Benzonase Analysis of β_A-globin LE and mutant 1 (m1) fragments

Fragments of the β^A-globin LE and mutant 1 (m1) promoter regions were prepared as described above. *Cla*I – *Hinf*I fragments were labeled with ^{32}P-γ-ATP and one of the labeled ends removed with *Pst*I. Thus, the forward strand of both the LE and m1 DNA fragments were labeled. The mutant m1 contains the radiolabeled sequence 5'...GGGCCGCGCTGTGCT...3'. Highlighted in light green are the mutated nucleotides. The mutation results in a sequence reduced from a CpG triplet to a CpG doublet. Benzonase digests were carried out on these fragments according to the standard protocol, with increasing amounts of the Benzonase enzyme.

The digests were purified as described above and run on an 8 % denaturing PAGE/urea gel. Figure 4.7 shows the results of this experiment with a highlighted region containing the CpG triplet. The Benzonase digests on m1 provide a similar pattern to the wild-type pattern. Comparisons of lane 2 for the methylated LE fragment to lane 2 for the methylated m1 fragment in figure 4.7 shows that Benzonase still shows hypersensitivity at the mutated triplet. The cleavage is not as pronounced as the wild-type LE fragment, but exists over the same region nonetheless. The methylated sites in m1 are still embedded within alternating base pairs, but instead of a row of the wild-type eight alternating bases, there are now five. Instead of three methylatable sites there are now two. The effect of Benzonase cutting is reduced by the loss of one methylatable CpG but not enough to eliminate a preferred cutting site. The banding pattern is less defined over the region especially at the central CpG in the triplet, which is the first CpG in the mutant. The first CpG in m1 is preceeded by a cytosine residue, which could detract from the A-form potential. I am basing this assumption on the crystallographic and structural data available to me (table 1.2). Interestingly, DNaseI data for this mutant (figure 4.5 *A*) shows that CpG methylation causes DNaseI hypersensitive sites in the fragment.

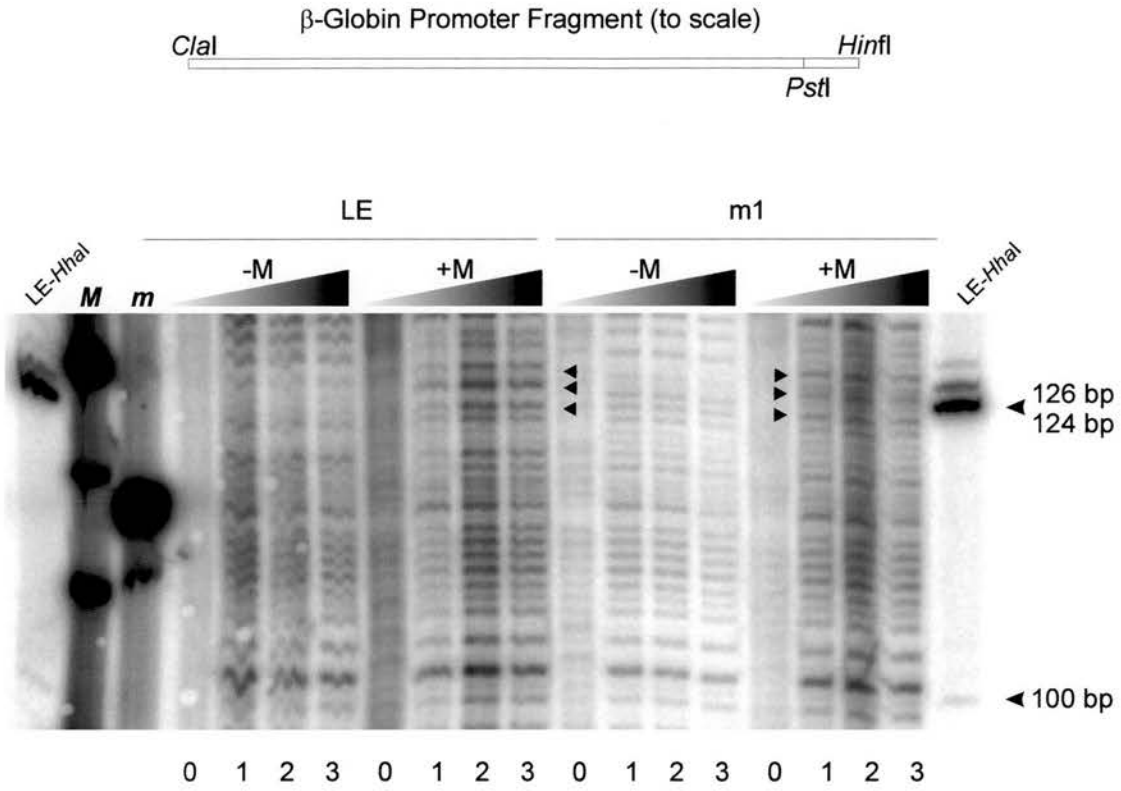


Figure 4.7 Benzonase digestion of wild-type LE fragments compared to mutant 1 (m1). Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet at 124 bp; ***M***, pBR322-*Bsu*RI digest marker; ***m***, pUC19-*Msp*I digest marker. Arrows indicate sites of methylation induced structural change.

Comparing lanes 2 and 3 shows that there is still methylation enhanced cleavage, but predominantly at one site and not at all three as displayed in lane 1. The corresponding nucleosome mapping data for this mutant shows that it is still capable of altering nucleosome positioning, when this CpG doublet is methylated (figure 4.6 *B*). Lanes 2 and 3 show that nucleosome 5A occupancy has been reduced at this position, but not eliminated as successfully as the wild-type LE fragment (lane 1) nor indeed as successfully as some of the other mutants.

It follows, that if CpG methylation in certain sequences and at certain positions can form A-form DNA, and a subset of these sites correlate with regions that have the ability to displace nucleosomes - then perhaps altering DNA tertiary structure is a feature that is used at other specific sites in the regulation of nucleosome positioning and potentially, in the regulation of gene expression. If this hypothesis is true, we would expect to see a correlation between A-DNA formation and loss of nucleosome positioning and *vice versa*, a correlation between the lack of A-DNA formation and loss of the ability to move nucleosomes. This may indeed be the case for m1. It is less likely to form A-DNA as it has less methylatable sites and less alternation of GC base pairs. It may partially form A-DNA, but this may not be enough to exclude nucleosome 5A completely from this region when methylated.

4.1.2.5 Benzonase Analysis of β^A-globin LE and mutant 2 (m2) fragments.

To further investigate the nature of the structure of the CpG triplet, and to correlate the nucleosome positioning data of Dr. C. Davey with my observations on DNA structure, fragments of the β^A-globin mutant no 2 (m2) were taken for comparison to the wild-type. The m2 fragments contain the sequence 5'...GGCGGCCGCTGTGCT...3'. Highlighted in yellow are the CG dinucleotides, in blue the alternating nucleotides and in light green, the mutated nucleotides. Again, these fragments were prepared as described above for LE and m1 forward strands and Benzonase analysis carried out as before. Figure 4.8 *A* shows the results of this analysis.

As can be seen from figure 4.8, the methylated LE fragment shows increased cutting in and around the CpG triplet when compared to the unmethylated LE fragment. This was used as a control reaction. In direct comparison, the

unmethylated m2 fragment behaves similarly to the unmethylated LE fragment over the region of the CpG triplet (figure 4.8 *A*). The pattern of digestion (lanes 1 & 2) in the methylated m2 fragment, does show increased cutting around the CpG triplet, but is not quite as defined as the LE fragment. This experiment was repeated producing the same result. A comparison was made of the m2(-M) fragment and the m2(+M) fragment using Aida™ software. The results of these scans are shown in figure 4.8 *B*. This scan is not as smooth as some of the others, but despite differences in intensity it is clear that there are structural changes present that are not present in the unmethylated fragment.

All of the available crystallographic data (Tippin, *et al.,* 1997), suggests that this sequence may also form A-DNA or at least partly do so. As can be seen from the m2 sequence, the mutant version is not too removed from the wild-type: (1) It still retains a purine start before the alternation of the first GC dinucleotides begins, this may contribute significantly to the promotion of A-form DNA, but (2) it has lost a CG dinucleotide pair. As this disrupts the alternation of GC dinucleotides, it may detract from the formation of A-DNA. (3) There are two methylatable sites in the sequence as opposed to three in the wild type. Although the methylatable CpGs are not adjacent to each other, they contain more alternation of base sequence around these CpGs than m1, which may explain why this mutant is slightly better at altering nucleosome positioning than m1 (see below). However on the whole, this mutant contains two methylatable sequences, like m1 so cannot be expected to behave too differently.

DNaseI analysis on m2 (figure 4.5 *A*) reveals that this sequence does show some enhanced cleavage at this position (lane 4 & 5), but it is less pronounced than LE and most similar to the behavior of m1. The corresponding nucleosome positioning data (figure 4.6 *B*) shows that nucleosomes can position at the unmethylated m2 sequence (lane 4). There is some loss of positioning at the methylated m2 sequence but it does not appear to be as effective as LE.

Overall, m2 may be expected to behave in a similar manner to the m1 sequence, with two me-CpGs. However, unlike the m1 sequence, the methylated CpGs are separated from each other. This results in a break in the alternation of CG dinucleotides. Therefore, although heightened cutting is observed around this

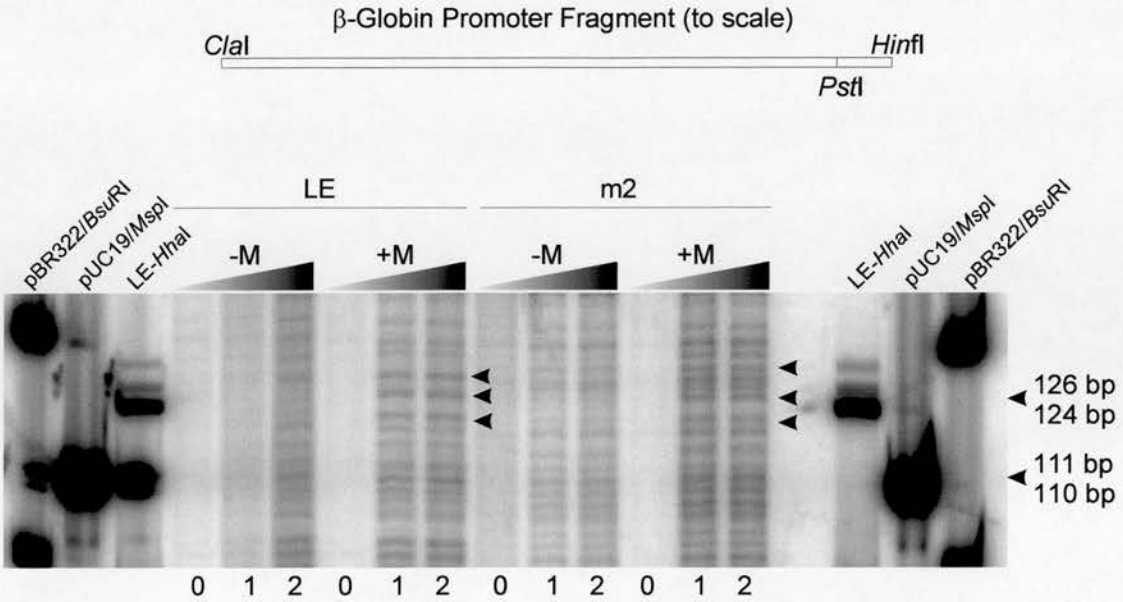


Figure 4.8 A Benzonase digestion of wild-type LE fragments compared to mutant 2 (m2). Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet. Arrows indicate regions where methylation inuced structural transitions are recognised by Benzonase.

B

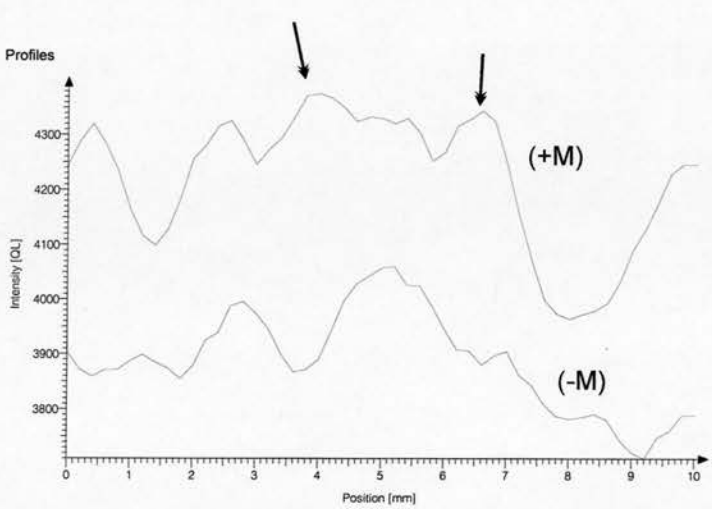


Figure 4.8 *B* Peak determination scan of CpG triplet region in the mutant m2. Peak determination scans were performed using Aida software (Fugifilm). Scans cover the *Hha*I site marked in the image above. Lane 2 of the unmethylated (-M) fragment is compared to lane 2 of the methylated fragment. The difference in starting intensity is due to the lower incorporation rate of the unmethylated m2 mutant. In this case the overall patterns should be compared. Increased cutting is seen in the methylated m2 fragment as compared to the methylated fragment. Arrows indicate regions of methylation enhanced cleavage as detected by Benzonase.

sequence, it is not so precise as the three discernible bands present in the methylated wild-type fragment or m1. In line with the hypothesis of A-DNA directed nucleosome movement at this position, the nucleosome 5A does move, but not as well or as completely as LE.

4.1.2.6 Benzonase Analysis of β^A-globin LE and mutant 3 (m3) fragments

This mutant contains the sequence 5'...GGCGCGGCCTGTGCT...3' and was prepared for analysis by the same methods as m1 and m2. The altered bases are highlighted in green and consist of an inversion of the third CpG dinucleotide in the CpG triplet. This results in a sequence that has two methylatable CpG dinucleotides in adjacent positions embedded in a row of five alternating nucleotides. Overall, the m3 mutant is most similar to wild-type in terms of alternation of base sequence around the methylation sites.

Figure 4.9 *A,* shows a Benzonase digest on the LE fragment and the m3 mutant in both unmethylated and methylated states. The area of interest is highlighted. As can be seen from this analysis the m3 fragment behaves very similarly to the LE mutant as expected from the sequence structure. That is, the m3 mutant still maintains a strong reactivity with the Benzonase assay in the CpG methylated state indicating that it still retains the propensity to form A-DNA. Comparisons of lane 2 in each digested fragment reveals that both LE and mutant 3 respond very similarly. An Aida trace scan (figure 4.9 *B*) of the region shows clear differences in the methylated and unmethylated m3 mutant fragments. The methylated m3 mutant has also been compared to the methylated LE fragment using an Aida™ trace scan (figure 4.9 *C*). Peaks in the methylated LE fragment are centrally located. Peaks in the methylated m3 fragment are located at the beginning of the scan at positions coinciding with the first methylatable CpG. From known crystal structures, the m3 mutant has the best chance of forming A-DNA in a similar capacity to the wild-type sequence. Although it has 2 methylatable CpGs and not 3, they are both adjacent to each other and preceeded by a purine residue. In both m1 and m2 mutants, only one of the two methylatable sites are preceeded by a purine residue, and although both methylated m1 and m2 show some Benzonase reactivity, neither show banding patterns as similar to the wild-type situation as m3, in terms of

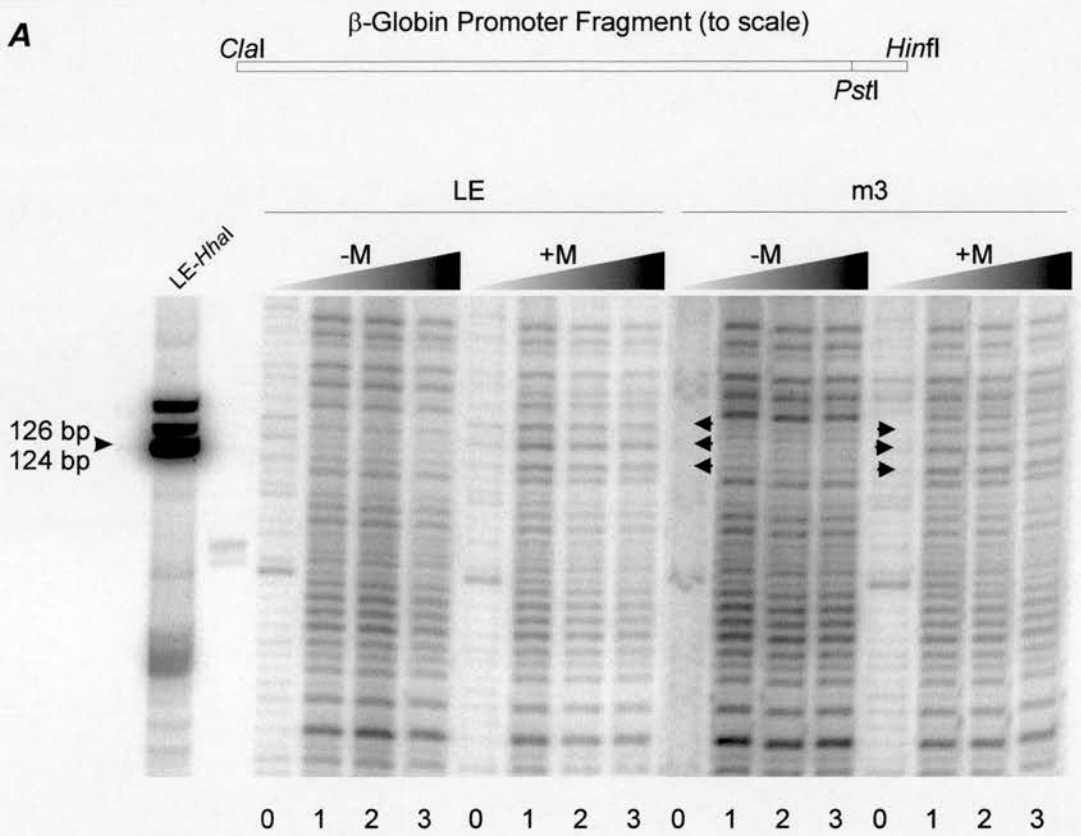
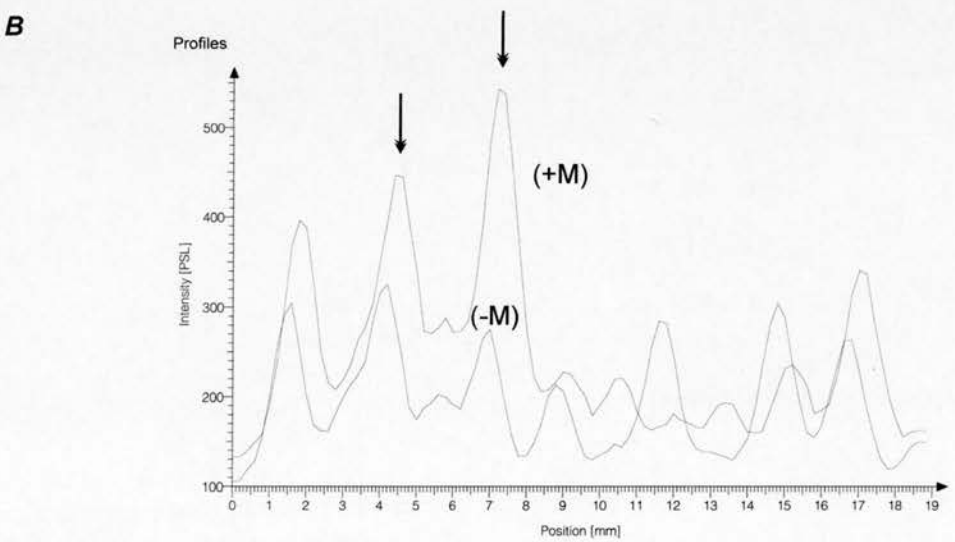


Figure 4.9 *A* Benzonase digestion of wild-type LE fragments compared to mutant 3 (m3). Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet. Arrowheads indicate methylation induced structural changes.



Figuew 4.9 *B* Peak determination scans of the CpG triplet region of mutant m3. The m3 mutant is here compared in unmethylated (-M) lane 3 of 4.9 A, and methylated (+M) lane 1 of 4.9 B. Methylation enhanced cleavage is shown by an arrow.

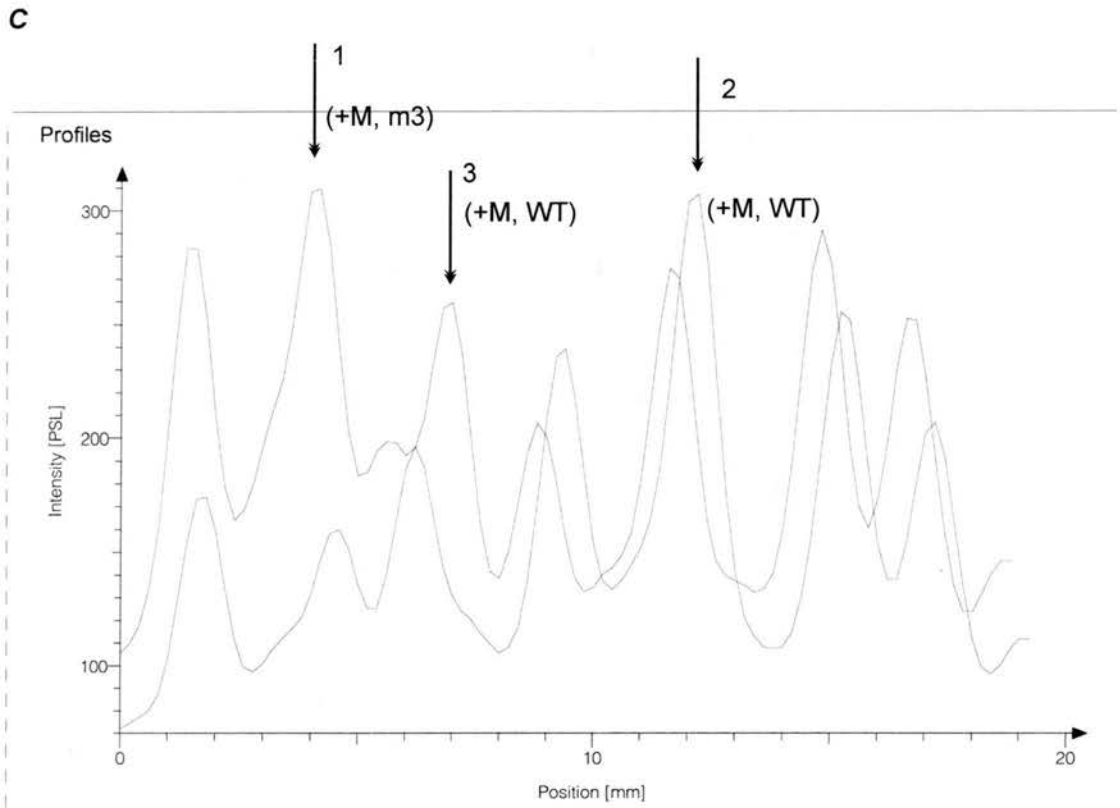


Figure 4.9 *C* Peak determination analysis of a region of the mutant m3. This image demonstrates the subtle differences shown in methylation (+M, WT) of the wild-type fragment and methylation of the mutant m3 fragment (+M, m3). In this case arrows indicate sites of methylation enhanced cutting. 1, refers to enhanced cleavage over the mutant m3 fragment. 2, refers to methylation enhanced cleavage over the methylated wild-type LE fragment. Peak 3 shows enhanced cleavage at the methylated LE fragment, that is not present on the methylated m3 fragment. Typically, the methylated LE fragment has more centrally located hypersensitive sites, probably reflecting the methylated triplet. The m3 fragment has hypersensitive sites at the beginning of the region, probably reflecting the methylated CpG doublet.

regular spacing between bands and methylation induced intensity. This is reflected in the nucleosome positioning data.

The DNaseI data (figure 4.5 *A*) shows heightened reactivity with methylated m3 (lane 6 & 7), but only over the methylated end of the fragment. Instead of three distinct bands (lane 1) there is one strong band corresponding to the methylated end, in an almost mirror image of what happens on m1 (lane 3) which of course has methylation at the opposite end of the mutated triplet than m3.

The mutant m3 is still capable of moving the position of nucleosome 5A away from this positioning site (figure 4.6 *B*, lane 6 & 7). This it does with better capabilities than both the mutant m1 and m2 fragments. This is consistent with the fact that it is most like the LE fragment is sequence organisation and Benzonase reactivity. It appears that the extent of movement of nucleosomes at this position may correlate well with the ability of this sequence to form A-DNA as detected by Benzonase.

4.1.2.7 Benzonase Analysis of β^A-globin mutant 23 (m23) fragments

The m23 mutant contains the altered sequence 5'...GGCGCCCGCTGTGCT...3'. The green nucleotide indicates the mutated base, the yellow blocks the methylatable sites and the blue nucleotides are alternating bases. Benzonase analysis was performed on radiolabeled fragments as described for the mutants above. The results of this experiment are shown in figure 4.10. Of note with this sequence is the fact that although there are two methylatable CpG sites present, they are (a) not adjacent to each other, and (b) separated by non-alternating cytosine dinucleotides. This type of arrangement is comparably less likely than either LE, m1, m2, m3 and m25 to form A-DNA at the second CpG. DNA beginning with the sequence d(CCCG) does not form A-DNA when methylated according to structural studies and instead forms Z-DNA as discussed earlier. It is possible that methylation over the first part of this sequence could serve to potentiate A-DNA whereas methylation over the second part of this sequence could influence the formation of Z-DNA. The Benzonase digest in figure 4.10 is unfortunately not very clear for the unmethylated region, but separate re-running of the unmethylated sample showed a digestion pattern for the unmethylated region similar to the methylated m23 fragment. It appears that although Benzonase cleaves the mutated region well, it has the same pattern of

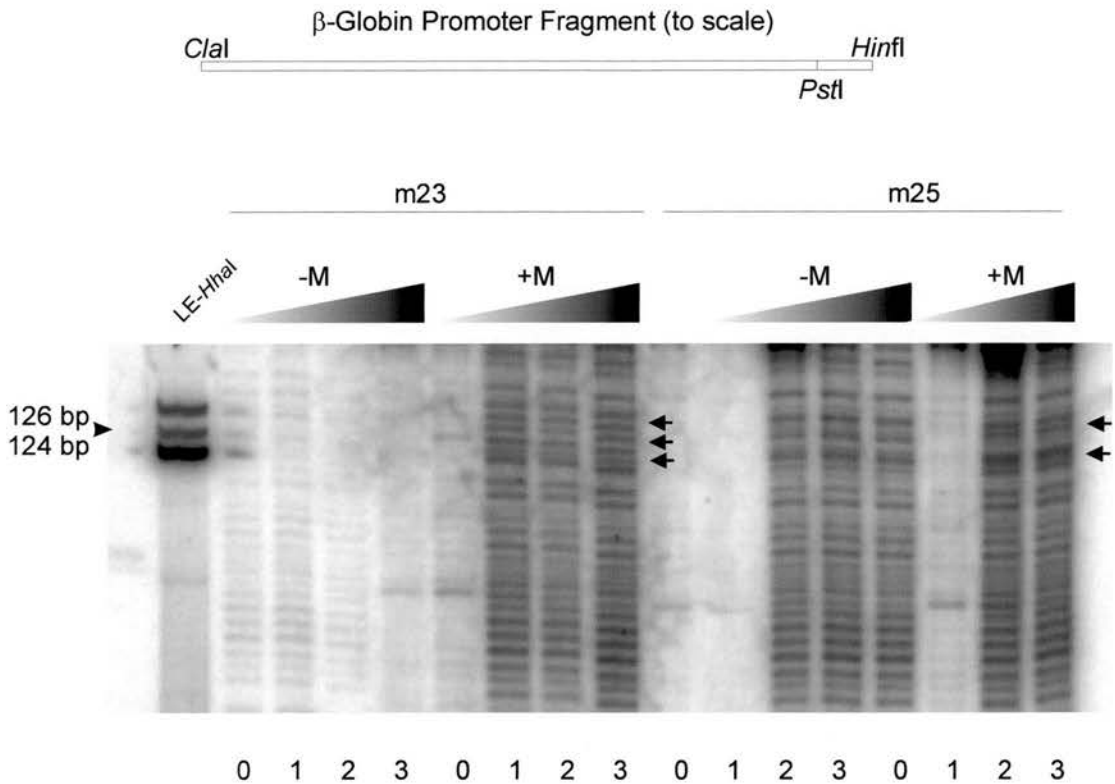


Figure 4.10 *A i* Benzonase digestion of mutant 23 (m23) and mutant 25 (m25) fragments. Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet. Arrowheads indicate regions of methylation induced structural change.

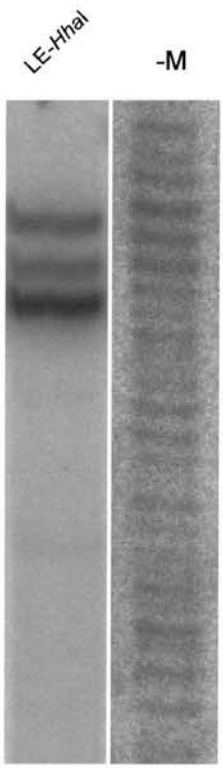


Figure 4.10 *A ii* Enhanced view of the m23 mutant. This image is supplied to show the unmethylated m23 mutant from figure 4.10 i above. The unmethylated m23 mutant lane was separated from the data and the contrast enhanced such that the banding pattern is more visible for this mutant. Lane 1 was chosen for this analysis

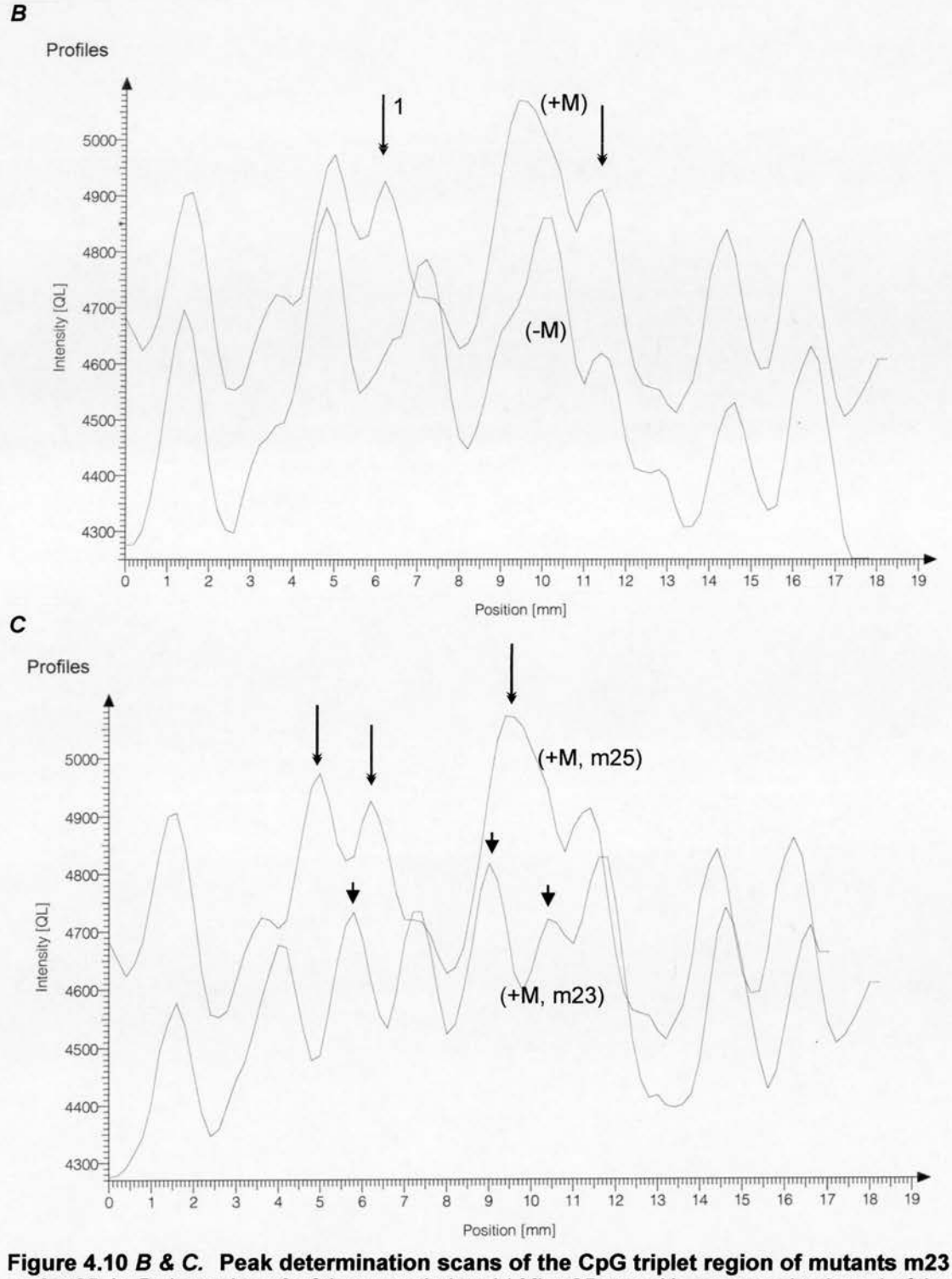


Figure 4.10 *B & C.* Peak determination scans of the CpG triplet region of mutants m23 and m25. In *B* above, lane 3 of the unmethylated (-M) m25 strand is compared to lane 1 of the methylated m25 (+M) strand. The intensities of each lane are different as is observed above, but the peak pattern is similar over the region. Methylation increases Benzonase reactivity at some regions that are marked with arrows. At these points a degree of band tightening is observed. In image *C,* the methylated m25 mutant is compared to lane 1 of the methylated m23 mutant. The m23 mutant seems to be quite regular in its banding pattern. However it does not appear to show any major enhancements at methylated sites as we observe in other mutants. Note that m23 does not shift nucleosome 5A in the β-globin gene fragment. It does not exhibit the same hotspots of reactivity as the m25 mutant. This may provide part of the explanation for why m23 is inefficient at nucleosome displacement. Notable differences are marked with thick arrows.

digestion as the unmethylated fragment (4.10 *A, ii*). So, although there are regular bands in the cleavage pattern, they appear to be the same regular bands as appear in the unmethylated fragment also. This could mean that methylation of m23 does not change the sequence structure significantly enough to induce an A-form structure. It is difficult to draw too many conclusions from m23 in terms of Benzonase analysis of the m23 mutant alone, but the notion that this sequence may have a different structure at the different methylation sites, based on the 'rules' for A-DNA formation established here, seems plausible when the nucleosome data is considered.

DNaseI data for mutant m23 (figure 4.5 *A*, lanes 8 & 9) shows two enhanced cleavage sites corresponding to the two methylated CpGs, confirming that there is some type of structural change at both of these positions. The nucleosome positioning data (figure 4.6 *B*) shows that this sequence can position nucleosomes well in the unmethylated state (lane 8). Unlike any of the other mutants, the positioning of nucleosome 5A is not moved by CpG methylation (lane 9). This is an interesting result, as from this analysis, it appears that despite two methylation sites on mutant 23, at the same positions as in the m2 mutant, methylation does not cause the nucleosome to move. The intervening sequence may be having an affect and as stated may be more likely to form Z-DNA. Negative supercoiling is known to promote Z-DNA formation and this is a feature of nucleosomal DNA. Therefore any part of the DNA sequence containing Z-DNA in this case may be incorporated well into nucleosomes. This may be one explanation for loss of movement of nucleosomal 5A when the m23 sequence is methylated.

As stated, Benzonase is reacting with both unmethylated and methylated sites in the m23 mutant. Despite the fact that the cleavage data for the unmethylated m23 is of low intensity, in my estimation there is no significant change in the pattern between the different methylated states for m23. The lack of a significant methylation induced structural change at m23 could therefore account for the lack of nucleosome displacement observed at the methylated m23 sequence. I conclude that for reasons discussed, the possibility exists that (a) the extent of A-DNA formation on methylated m23 may be little or none, (b) the change in structure is thus too short to displace nucleosome positioning effectively and possibly (c) the second

methylation site may be promoting Z-DNA formation and as a result is incorporated well into nucleosomes, or at least does not repel them at this position.

4.1.2.8 Benzonase Analysis of β^A-globin mutant 25 (m25) fragments

The m25 mutant contains the altered sequence 5'...GGCGGGCGCTGTGCT...3'. This mutant is similar to the m23 mutant in the sense that there are still only two non-adjacent methylatable CpG sites. However, they are separated by two central G residues as opposed to two central C residues. Also, there is alternation of base sequence over both of these CpG sites, unlike the m23 sequence which has alternation over the first CpG only. Each methylatable site is preceeded by a Guanine residue in an alternation known to promote A-form DNA. This would imply that perhaps the m25 mutant sequence can form A-DNA over both regions unlike m23 above.

Benzonase analysis of the m25 sequence is shown in figure 4.10, alongside the m23 data. As predicted from the DNA sequence, known crystallographic data and the data shown here, the methylated m25 sequence reacts well with the Benzonase enzyme over the mutated region and shows preferred cut sites in the methylated fragment (compare lane 3 –M and lane 2 +M).

The DNaseI data (figure 4.5 *A*), for m25 confirms methyl-CpG induced structural alternations. It shows enhanced cleavage at the methylated triplet region. The enhancement is not as strong as the methylated LE fragment at any of these positions and looks most similar to the m2 mutant. It does not have the same regularity of spacing as the LE bands, nor the same intensity of cleavage. In fact m25 is almost identical to m2 except for one nucleotide. The nucleosome positioning data (figure 4.6 *B*) shows that the unmethylated m25 sequence can position nucleosomes (lane 10) but not as effectively as LE. From what positioning is available on this sequence, it appears that m25 can displace its positioning with comparable effects to m1 and m2.

An Aida™ scan of the CpG triplet region for Benzonase digestion of the unmethylated and methylated m25 fragment is shown in figure 4.10 *B* with changes in banding pattern highlighted. In 4.10 *C*, the methylated m23 mutant, which cannot displace nucleosome 5A, is compared to the methylated m25 mutant for band

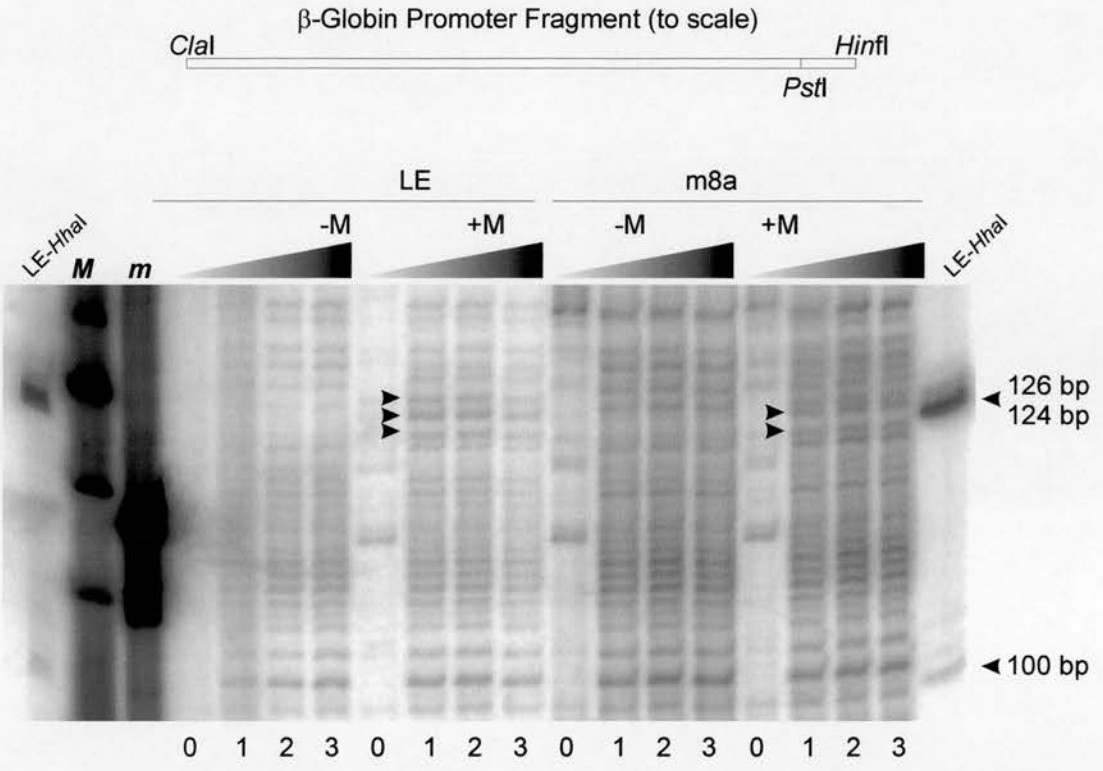


Figure 4.11 *A* **Benzonase digest of wild-type LE fragments compared to mutant 8a (m8a).** Benzonase digestion was performed on *Sss*I methylated (+M) and unmethylated (-M) β^A-globin promoter fragments. LE-*Hha*I, β-globin promoter fragment digested with *Hha*I, which represents the site of the CpG triplet; ***M***, pBR322-*Bsu*RI digest marker; ***m***, pUC19-*Msp*I digest marker. Arrowheads indicated regions of methylation induced structural change.

B

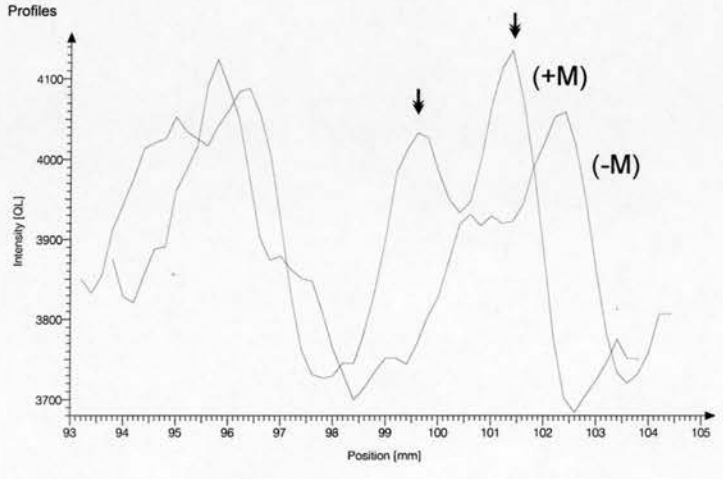


Figure 4.11 ***B*** **Peak determination scans of the m8a mutant.** Peak determination scans were performed for unmethylated (-M) from lane 2 of 4.11 *A* and methylated (+M) from lane 1 of fig. 4.11 *A*, DNA fragments are shown for the m8a mutant. The m8a mutant shows a lot of similarity over its general banding pattern, but there are some marked differences. These are marked with arrows above the methylation induced peaks.

distribution. The methylated m23 mutant appears to display an even enough range in band distribution and intensity values. Differences in peak positions are indicated.

4.1.2.9 Benzonase Analysis of β^A-globin LE and mutant 8a (m8a) fragments

The m8a fragment contains the sequence 5'...GGCGGCCGCTGTGCG...3'. This mutant is interesting as it contains only two of the wild-type methylatable CpG sites. These are in the same organization as the m2 mutant. The difference is that there is an additional mutation 7 bp downstream of the CpG triplet region in m8a, producing a third methylatable CpG site. As seen from figure 4.11, the methylated m8a mutant sequence reacts more strongly with Benzonase than the unmethylated version as expected. Like m2 (figure 4.9), the Benzonase digestion pattern is enhanced at the methylated sites. Also like m2, the pattern is not as defined as in LE. An Aida™ scan of the region (figure 4.11 *B*) shows that the unmethylated and methylated m2 banding patterns are very similar but the methylated m2 pattern is shifted to the left in a manner that coincides with enhanced additional DNA bands in the sequence.

The DNaseI data (figure 4.5 *A*, lanes 12 & 13) shows enhanced cleavage over the methylated cytosines. Unsurprisingly, the nucleosome positioning data for m8a (figure 4.6 *B*, lanes 14 & 15) shows a pattern almost identical to m2 (lanes 4 & 5), which is also very similar to the m25 mutant. Nucleosome positioning is very good in the unmethylated fragment and it is removed on methylation with comparable efficiency as the m2 or m25 mutant. Together these mutants are the most similar in sequence construction of all of the mutants and behave similarly as shown here.

4.2 Summary

In summary, a number of observations and conclusions can be made.

(1) That CpG methylation of the CpG triplet in the β^A-globin LE fragment causes a significant structural alteration. This alteration has changed the structure so much that it now becomes a preferred cleavage site for the DNA digesting enzyme, Benzonase. As (a) Benzonase is known to have a recognition and cleavage preference for A-form DNA, and (b) methylated DNA in this sequence context is known to form A-DNA from crystallographic data, it is reasonable to conclude that

by this *in vitro* assay, the methylated CpG triplet is forming a structure consistent with A-DNA.

(2) The enhanced cleavage seen at the CpG triplet on the forward strand is also seen on the reverse strand. This corroborates the hypothesis that this structure is forming A-DNA and not some altered secondary structure in this sequence that makes DNA more accessible to the Benzonase enzyme.

(3) There are other CpG methylatable sites in the β^A–globin fragment being analysed. None of these sites promote the formation of A-DNA. From analyzing these sites and the surrounding sequence context, and comparing these to known crystal structures, it appears that some degree of alternation of methylatable CG sites is required for Benzonase hypersensitivity. The formation of A-DNA is also known to be affected by the extent of the alternation of dinucleotides. This is clearly seen from the analysis of the various mutants. Typically the best Benzonase reactive sequence is the wild-type LE fragment with three methylatable CpGs. Benzonase banding patterns are regularly spaced over this region. The methylated mutants mostly show some reactivity with Benzonase, but the pattern is never the same or as regularly spaced as the LE fragment. All show some DNaseI reactivity as expected. Nucleosome displacement efficiency also changes considerably over the mutant series with the best A-forming sequences (LE, m3, then m2 and m25) showing the best nucleosome displacement and m23 showing no significant nucleosome displacement.

(4) Finally, the structural observations observed may have some direct relevance to gene regulation. As described in section 4.1.1, this region of the β^A-globin gene has been analysed previously. Davey *et al.*, (1997) have shown that this CpG triplet when methylated disrupts the positioning of a specific nucleosome. Z-DNA analysis of the structure showed that the triplet could form A-DNA, but only under negative superhelical stress. The CpG triplet is part of the promoter region of the β^A-globin gene and may be a site of some importance in transcriptional regulation.

Chapter 5: Results (III)

5.1 Bandshift experiments with purified methyl binding proteins.

I have established that CpG methylated DNA can induce dramatic changes in DNA structure in certain sequence contexts. This structure may be an A-DNA structure as detected by Benzonase. The sequences that react the best with Benzonase are those of the type d(GCGCGC), when methylated. This coincides with structural data from a variety of sources. Different structures may form in DNA due to different methyl cytosine densities in a specific sequence context. This may be part of the key to understanding how different methyl-binding proteins recognise and bind to different methylated CpGs. For example the formation of A-DNA at methylated $d(GC)_n$ sequences may be a feature recognised by some methyl-binding proteins, but perhaps may be feature refractory to other methyl-binding proteins. To begin an investigation into this hypothesis a series of Bandshift assays were performed. For these a series of proteins were purified and a range of DNA oligomers specifically designed for this purpose.

5.1.1 Design of oligos for Bandshift analysis.

A series of oligos were designed on the basis of NMR and crystallography studies. Oligomers were chosen for their documented ability to form A-DNA or B-DNA. Additionally a sequence containing a region corresponding to the β^A-globin gene CpG triplet sequence (see chapter 4) is used. It is known to change structure when methylated, showing structural features consistent with the formation of A-DNA, as detected by Benzonase (this thesis). Table 5.1 below summarises this information. All oligos were made double-stranded by annealing each one with their complementary oligo as described in Chapter 2.

5.1.2 Purification of the MBD domain of MeCP2.

The ≈10 kDa MBD domain of MeCP2 was purified by two methods. Firstly from Glutathione (GST)-fusion plasmid expression in *E. coli,* and secondly from a histidine (HIS)-tagged fusion plasmid. The HIS tagged protein was found to be the

most active, devoid of proteolytic digestion and the most stable in solution. The purification results of this protein are shown below.

The full protocol for MBD purification is described in Chapter 2. In brief, a pET6his plasmid containing the MBD domain of MeCP2 was transformed into BL21(DE3)pLysS cells. The cells were grown and protein expression induced with IPTG. The cells were harvested and the cell protein extracts purified. Samples were dialysed to remove urea and SDS-PAGE gels run at each stage of the initial purification process (data not shown). Combined samples were applied to a Fractogel EMD-SO_3-650 cation exchange column to remove *E. coli* proteins and to concentrate the extracts. The bound protein was washed with increasing NaCl concentrations. The results of this experiment are shown in figure 5.1 *A* and *B*. Lanes 8 to 12 were combined, diluted and applied to a Ni^{2+} charged affinity column. The column was washed with increasing concentrations of L-histidine. Samples were taken and applied to an SDS-PAGE gel, the results of which are shown in figure 5.2 *A, B* and *C*.

Oligo	DNA sequence	Type
B-dnaU	5'-CGCATATATGCGCGCATATATGCG	B-DNA
A-dnaU	5'-GTACGTACGTACGTA-3'	A-DNA
B-gloUF	5'-GCCTGGCGCGCGCTGTGCT-3'	Z-DNA
B-gloUR	5'-AGCACAGCGCGCGCCAGGC-3'	Z-DNA
A-dnaM	5'-GTAmCGTAmCGTAC-3'	----------
B-gloMF	5'-GCCTGGmCGmCGmCGCTGTGCT-3'	A-DNA
B-gloMR	5'-AGCACAGmCGmCGmCGCCAGGC-3'	A-DNA

Table 5.1 DNA oligos used for Bandshift analysis. The oligos were designed for use according to their structural characteristics. mC, methylation at 5' position of cytosine. BdnaU, self-annealing B-DNA structural oligo. A-dnaU, self-annealing A-DNA structural oligo, Unmethylated. A-dnaM, is the same as A-dnaU but it is methylated. It is not clear what structure this may form when methylated. B-gloUF and B-gloUR anneal with each other to form an unmethylated dsDNA oligomer. B-gloMF and B-gloMR are methylated versions of the same fragment. All four B-glo oligomers correspond to the CpG triplet region of the β^A-globin gene (see chapter 4).

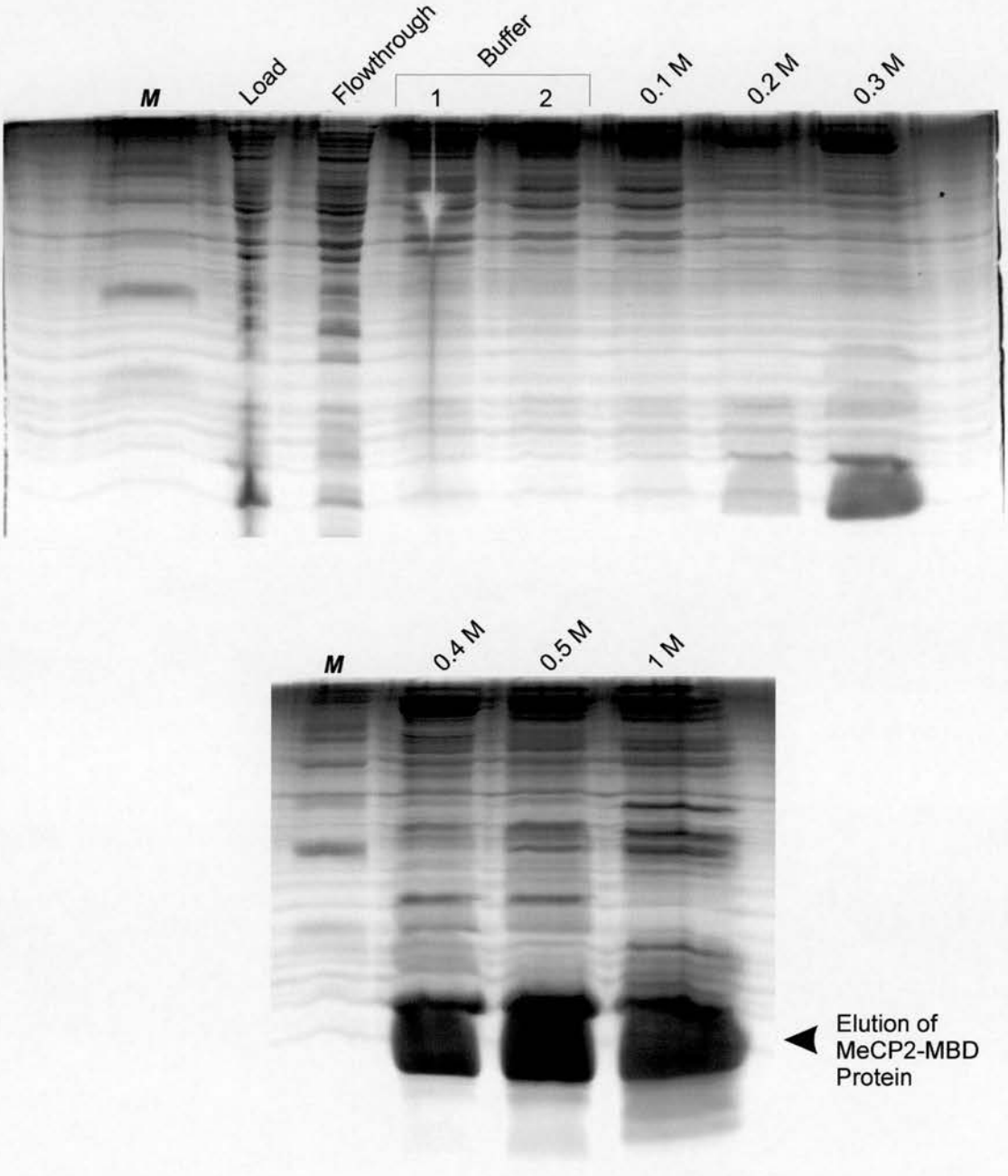


Figure 5.1 Purification of the MBD domain of MeCP2 using a cation exchange column. The *E. coli* expressed MBD protein was loaded onto an EMD SO_3-(S), (M) column (Merck) in Buffer A, to separate it from bacterial proteins. The MBD protein was eluted with increasing concentrations of NaCl in Buffer A. Fractions were collected as 1 to 2 ml samples. **1,2** refers to washes of the bound protein with Buffer A. **M,** Protein Marker, broad range (Bio-Rad). The MBD protein is eluted maximally between 0.4 M and 1 M NaCl. The samples from 0.4 M, 0.5 M and 1 M eluents were combined and dialysed to reduce the NaCl concentration. The combined sample was then further purified (fig. 5.2).

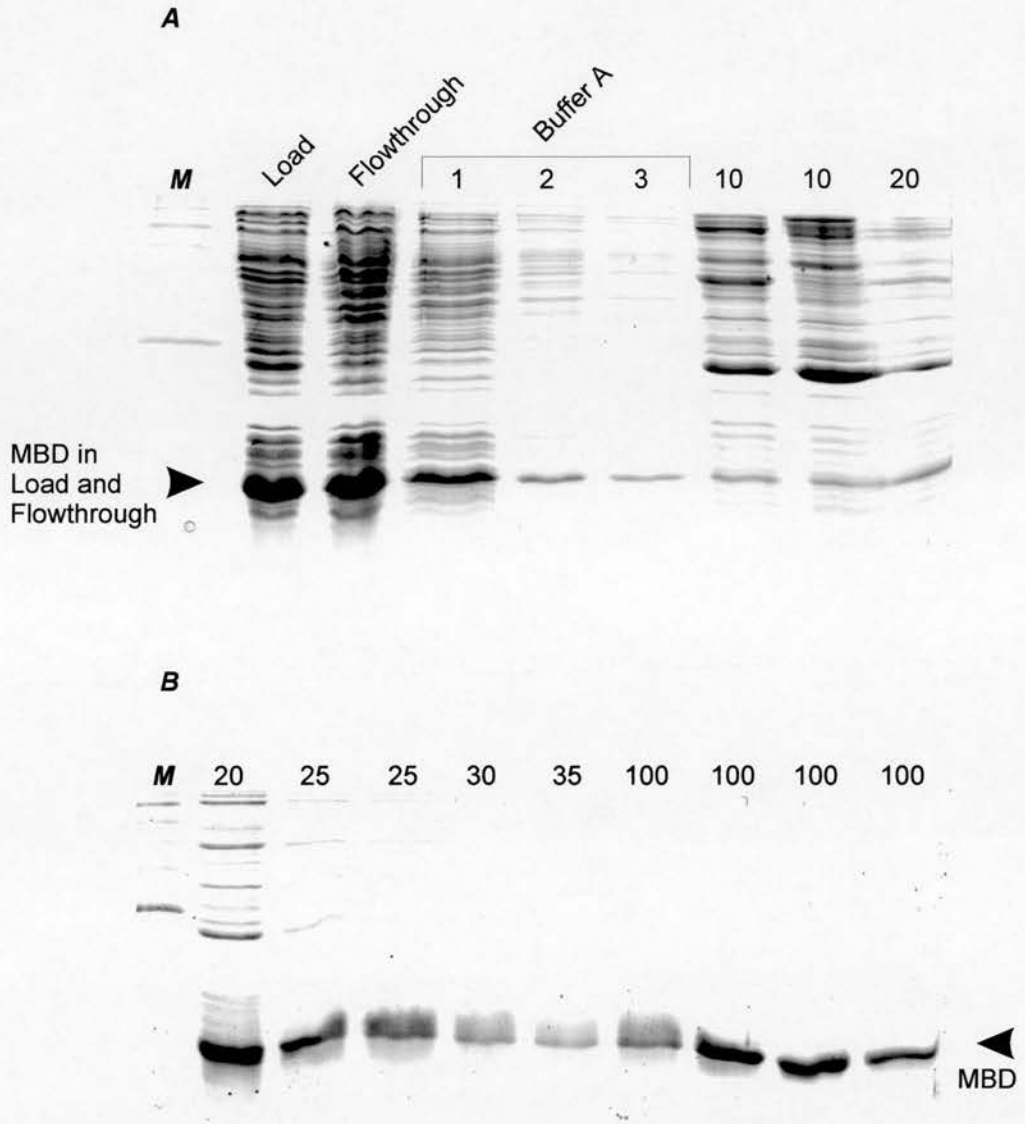


Figure 5.2 Purification of the MBD domain from MeCP2 using a Nickel affinity column. MBD purifications from an EMD-SO_3- cation exchange affinity column were applied to a Nickel charged EMD chelate 650 affinity gel in Buffer A solution. The column was washed three times with Buffer A, indicated by **1, 2, 3** in fig. ***A*** above. Low concentrations of L-histidine were used to elute unbound proteins by competing for Ni^{2+} binding. Concentrations of L-histidine are indicated by 10, 20 up to 100 mM. ***B,*** above shows the eluted MBD protein at high concentrations of L-histidine. This is indicated by a filled arrowhead. ***M,*** Broad range protein marker (Bio-Rad). Note that there is still some MBD protein in the load and flowthrough indicating that the concentration of MBD in the load was too high to completely bind to the Nickel charged column.

5.1.3 Bandshift experiments with the purified MBD domain of MeCP2.

The purified MBD domain of the MeCP2 protein is used in this analysis. The role of MeCP2 and its characteristics are discussed in detail in Chapter 1. The aim of these experiments was to investigate what types of sequence the MBD domain can recognise. It is known that MeCP2 can recognise a single methyl CpG site and that the MBD is responsible for this interaction (Meehan *et al.,* 1992, Nan *et al.,* 1993). Is this dependent on the sequence surrounding the CpG step? Is it the A-form structure of CpG methylated DNA that the MBD recognises and binds to? Can the MBD therefore recognise unmethylated sequences that potentiate A-form DNA? These questions were addressed by these initial bandshift experiments. From the DNA structural data obtained here it is clear that CpG methylation can induce A-type structures. However, the A-form structure is not seen at singly methylated CpG sites in my analyses if they do not occur within alternating $d(GC)_n$ sequences. In most alternating $d(GC)_n$ sequences in my analyses, the CpG steps are all usually methylated by *Sss*I methylase. Therefore the situation of a singly methylated CpG in a $d(GC)_n$ stretch does not arise in many of the Benzonase digests (Chapters 3 & 4). However the crystal structure of $d(Gm^5CGCGC)$ resolves into an A-form structure (Mooers *et al.,* 1995) and it is therefore possible that the MBD can recognise A-form DNA. The solution structure of MeCP2 does not indicate that the protein recognises the A-DNA conformation but does indicate that the change in position of cytosine induced by CpG methylation provides a structure that fits into the binding site of MeCP2 (Wakefield *et al.,* 1999). The oligos listed in table 5.1 were used for Bandshift analysis with the purified MBD protein.

5.1.3.1 The MeCP2 MBD and the A-dnaM oligo.

A bandshift experiment with the A-dnaM oligo and the purified MBD was performed with increasing concentrations of protein. The A-dnaM oligo was labelled on one strand with γ-^{32}P-ATP and annealed to unlabelled copies of itself. The oligomer was incubated with increasing concentrations of MBD in 1 X bandshift buffer on ice, and the samples run on a 5% PAGE gel. More details can be found in Chapter 2. The results of this experiment are shown in figure 5.3. As can be seen, lane 0 contains no MBD and does not show any shift in position of the A-dnaM

oligo. Increasing concentrations of MBD are used in lanes 1 to 9. As can be seen, there is a specific bandshift with this oligo as indicated in figure 5.3. This sequence has two methylated CpGs that are not adjacent to each other, confirming that the MBD alone is capable of recognising a single methylated CpG.

5.1.3.2 The MeCP2 MBD and the A-dnaU oligo.

A bandshift experiment was carried out with the MBD protein and the A-dnaU oligo in the same manner as described above. The results are shown in figure 5.4. Increasing amounts of the MBD protein shows that it does seem to recognise the unmethylated oligo at higher concentrations of the MBD. This interaction is much weaker than that observed with CpG methylated DNA. The bandshift that occurs is not a smear however, but a discrete complex. This may indicate that the MBD is positioned at a particular region on the DNA.

5.1.3.3 The MeCP2 MBD and the A-dnaU and A-dnaM oligos with competitor.

A bandshift experiment with the MeCP2 MBD and the A-dnaU and A-dnaM oligos was repeated in the presence of competitor DNA. The competitor used was *E. coli* DNA as the traditional salmon sperm is CpG methylated and would interfere with the results. The oligos were incubated with equal concentrations of the MBD and increasing concentrations of competitor DNA. The results are shown in figure 5.5. The results show that even in the presence of competitor DNA, the MBD still binds to the A-dnaM oligo specifically with no observable interaction with the A-dnaU oligo. Therefore the weak interaction seen with the A-dnaU oligo in figure 5.4 is now reduced by the presence of non-specific competitor DNA.

5.1.3.4 The MeCP2 MBD and the B-dnaU oligo.

Although the MBD binds only minimally to the unmethylated A-form oligo, a comparison was made to a typical B-form oligo, B-dnaU. This is shown in figure 5.6. It appears that the MBD can react with a typical B-form molecule, but the complex formed appears to be a smear of weak interactions and not a discrete complex. To do a side-by-side comparison, the double-stranded A-dnaU and B-dnaU oligos were again independently incubated with increasing amounts of MBD protein

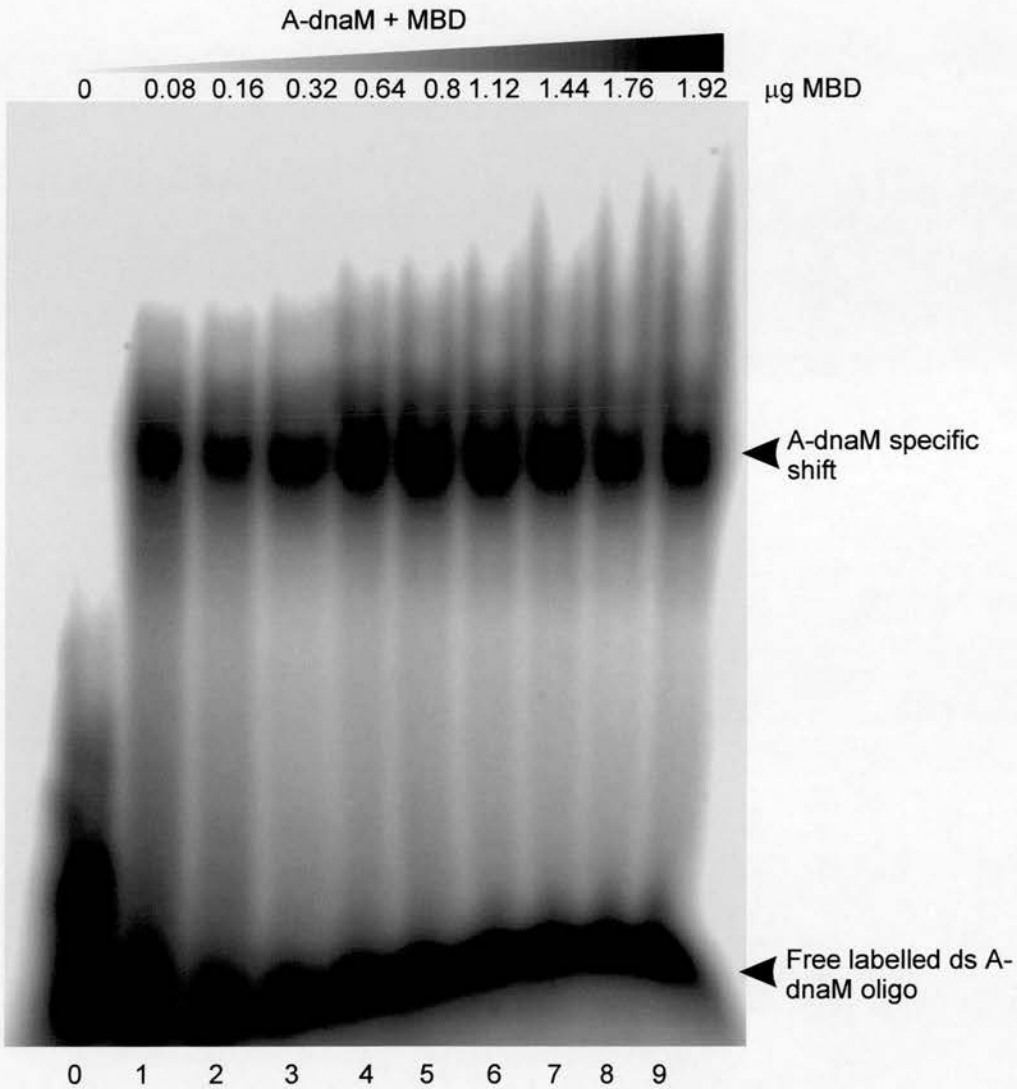


Figure 5.3 Bandshift experiment with purified MBD protein and A-dnaM oligomer. Double stranded and radiolabelled A-dnaM oligomers were incubated with increasing concentrations of MBD protein. Free DNA is indicated above by a closed arrowhead as are bound oligomers. Lane 0 is a control lane lacking the MBD protein. Samples were run on a 5 % polyacrylamide gel.

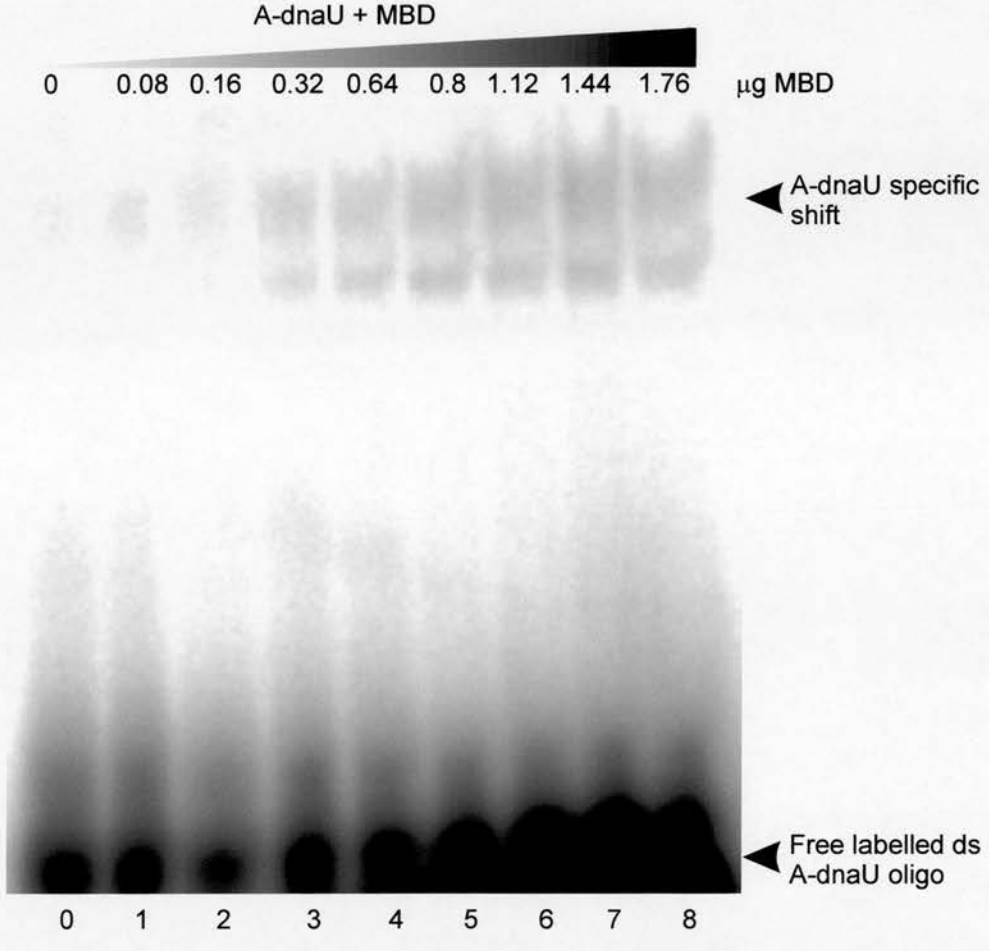


Figure 5.4 Bandshift analysis of the purified MBD domain of MeCP2 with the A-dnaU oligomer. The double-stranded A-dnaU oligomer was incubated with increasing concentrations of the purified MBD protein as indicated above. Arrowheads indicate the positions of free DNA and MBD bound oligomer. Lane 0 is a control lane lacking the MBD protein. Samples were run on a 5 % Polyacrylamide gel.

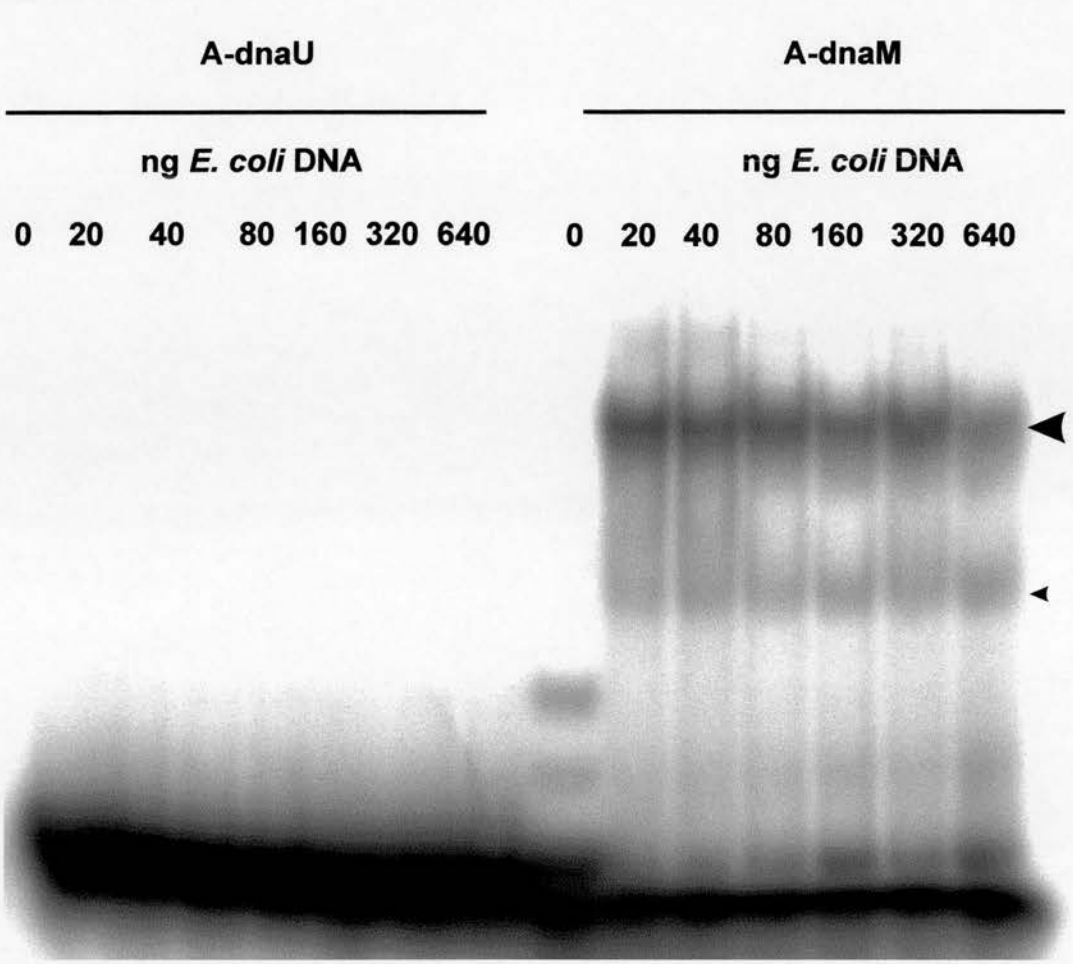


Figure 5.5 Bandshift analysis of purified MBD protein with the A-dnaU and A-dnaM oligomers in the presence of competitor DNA. Double-stranded labelled oligomers were incubated with a constant amount of MBD protein approximating 1 ug of protein per lane. Increasing amounts of non-specific, non-methylated *E. coli* DNA was added to each lane except for the 0 lane control in both cases, which contained no MBD and no competitor DNA. A-dnaM specific bandshifts are indicated with a filled arrowhead. Samples were run on a 5 % polyacrylamide gel. The presence of two bands may indicate the binding of one or two MBD occupancy sites on the CpG methylated DNA.

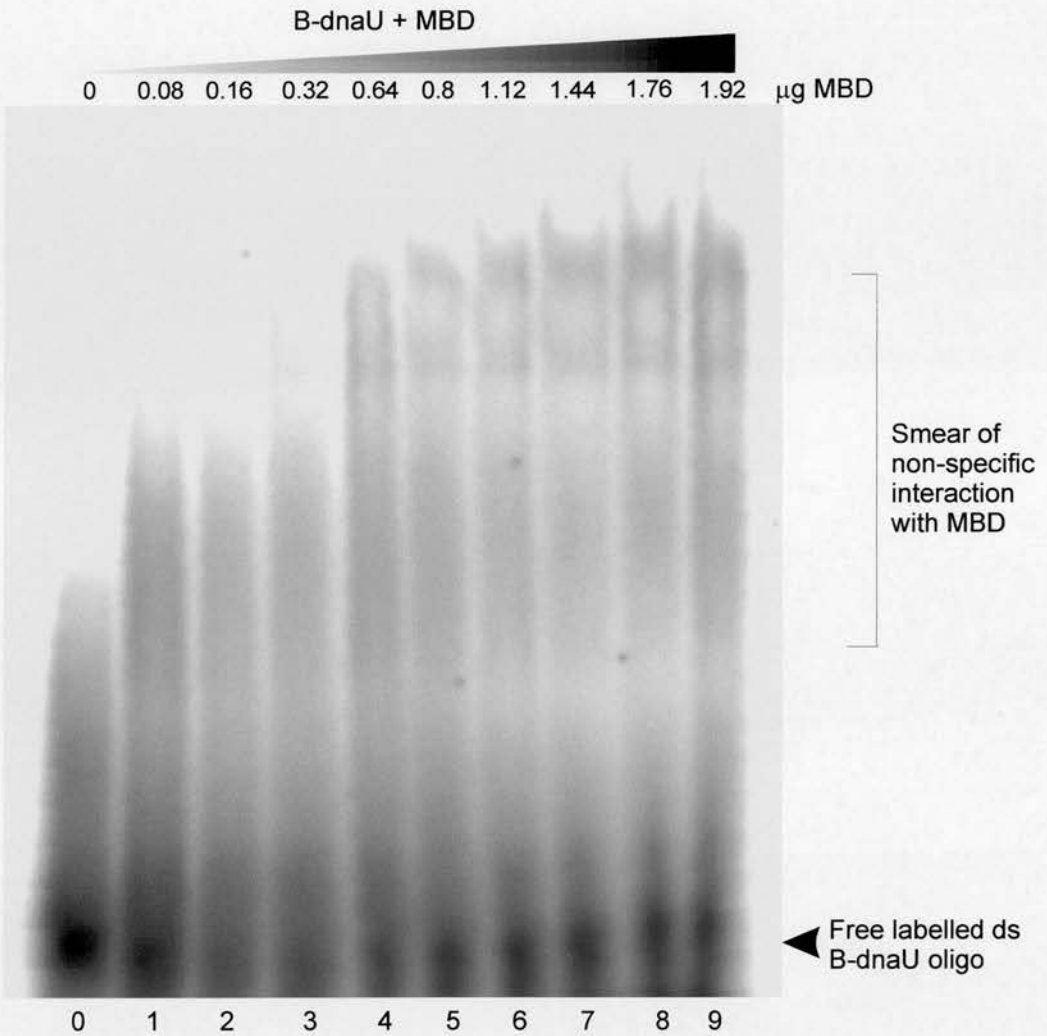


Figure 5.6 Bandshift analysis of purified MBD protein with the B-dnaU oligomer. Double-stranded DNA consisting of the radiolabelled B-dnaU oligomers were incubated with increasing concentrations of the purified MBD protein. The boxed region indicates an area of non-specific interaction with the B-dnaU oligomer. Free DNA is indicated by an open arrowhead. Samples were run on a 5 % polyacrylamide gel.

and run on the same gel. The results are shown in figure 5.7. It appears that the A-dnaU oligo reacts better than the B-dnaU oligo with MBD. This result warrants further analysis as it is a preliminary result. There is also no competitor DNA used in this analysis.

5.1.3.5 The MeCP2 MBD and the B-gloU and B-gloM oligomers.

To assess whether the MBD can bind to a methylated CpG string of dinucleotides as well as it does to a single methylated CpG, I have taken a region covering the CpG triplet region of the β^A-globin gene responsible for nucleosome displacement. It is probable that methyl-binding proteins are involved in the regulation of β^A-globin gene expression as described in chapter 4. To assess if the MBD can recognise and bind to this sequence, both B-gloU and B-gloM oligomers were incubated with increasing concentrations of MBD. The results are shown in figure 5.8. and figure 5.9. Clearly, the MBD can recognise the methylated CpG triplet specifically. There may be some non-specific contacts to the unmethylated CpG triplet, represented as a smear on the gel, but there is no specific complex formed.

5.1.3.6 Bandshift experiments with the purified xMBD1 protein.

To assess the sequence specificity of other MBD containing proteins on interaction with methylated DNA, the *Xenopus* MBD1 protein was used. This was provided by Dr. I. Stancheva who obtained the sequence by DNA cloning from a *Xenopus* λ-Zap library. The sequence was cloned into a pGEX-3X-GST fusion vector, expressed in *E. coli* and purified. The human MBD1 solution structure in complex with DNA has been solved (Ohki *et al.,* 2001). It has a very similar structure to MeCP2 across the whole structure, including the MBD. The details of hMBD1 are discussed in Chapter 1. The hMBD1 binds a singly methylated CpG dinucleotide in a similar manner to MeCP2. The xMBD1 protein can also bind a single methyl CpG. A comparison of the protein sequence of both is shown in figure 5.10 *A.* Bandshift experiments using the purified xMBD1 with the A-dnaU and A-dnaM double-stranded oligos is shown in figure 5.10 *B.* It is clear that xMBD1 can interact with the methylated oligo, but it also appears to interact well with the unmethylated oligo. This may be for several reasons. Perhaps there is not enough xMBD1 to provide a distinct interaction in the

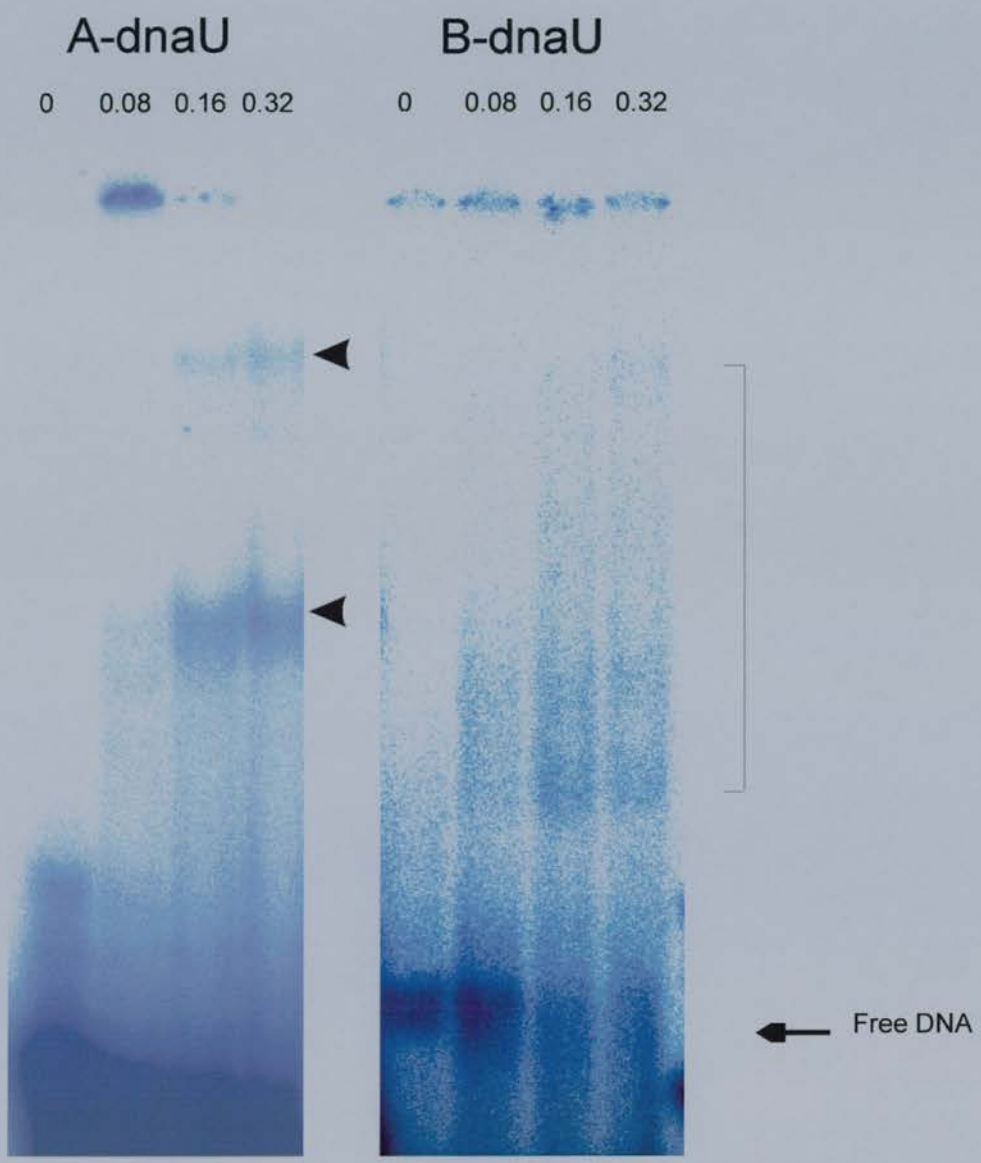


Figure 5.7 Bandshift analysis of the purified MBD domain of MeCP2 with the A-dnaU and B-dnaU oligomers. The double-stranded A-dnaU and B-dnaU oligomers were incubated with increasing concentrations of purified MBD protein as indicated. The boxed region indicates a region of non-specific interactions with the B-dnaU oligomer. The arrows indicate discrete bands representing MBD interactions with the A-dnaU oligomer at high concentrations of the MBD protein. Samples were run on a 5 % polyacrylamide gel.

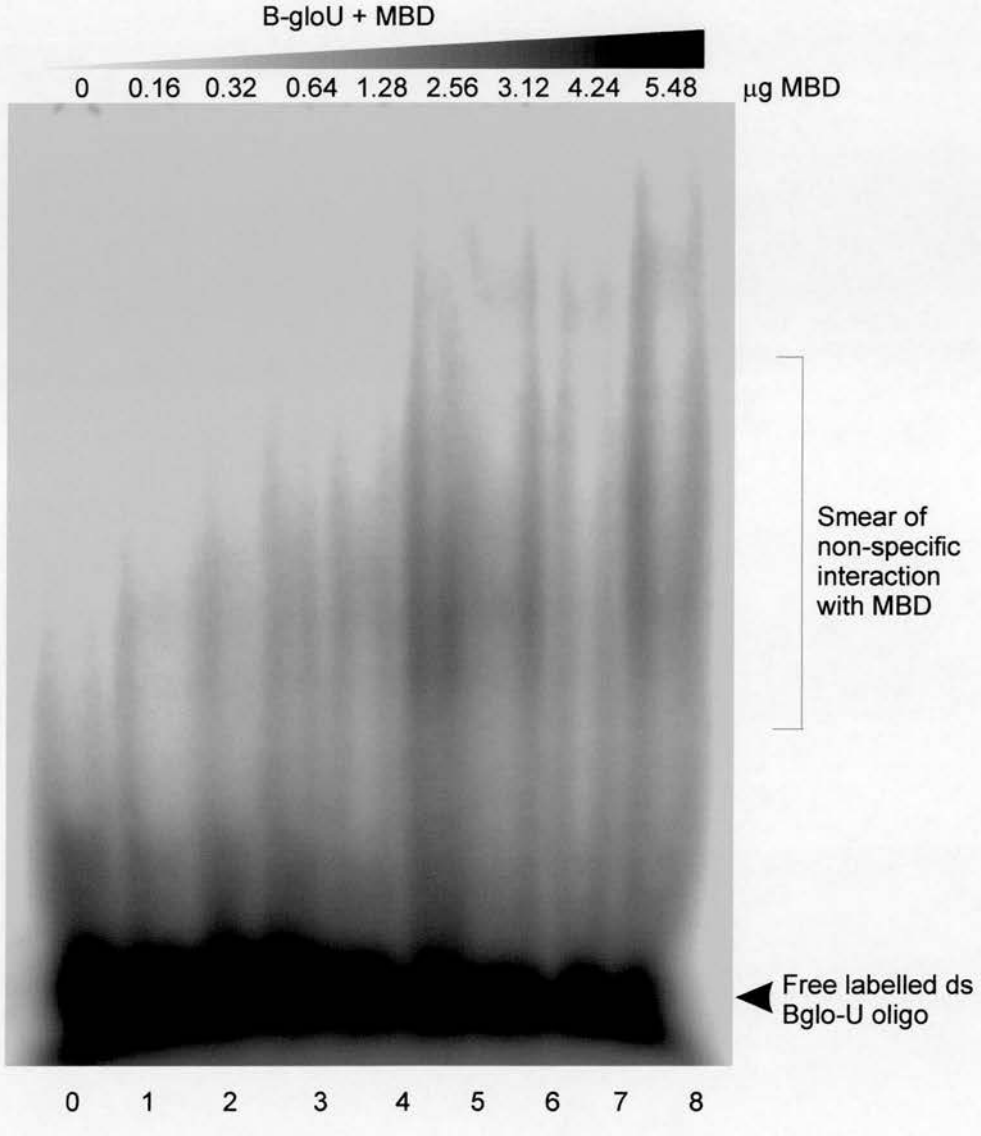


Figure 5.8 Bandshift analysis of the purified MBD domain of MeCP2 with the B-gloU oligomer. Double-stranded B-gloU oligomer was incubated with increasing concentrations of the MBD protein as indicated. Lane 0 is a control lane containing no MBD protein. Free DNA is indicated by a filled arrowhead. The bracketed region shows a non-specific smeared interaction with the B-gloU oligomer. Samples were run on a 5 % polyacrylamide gel.

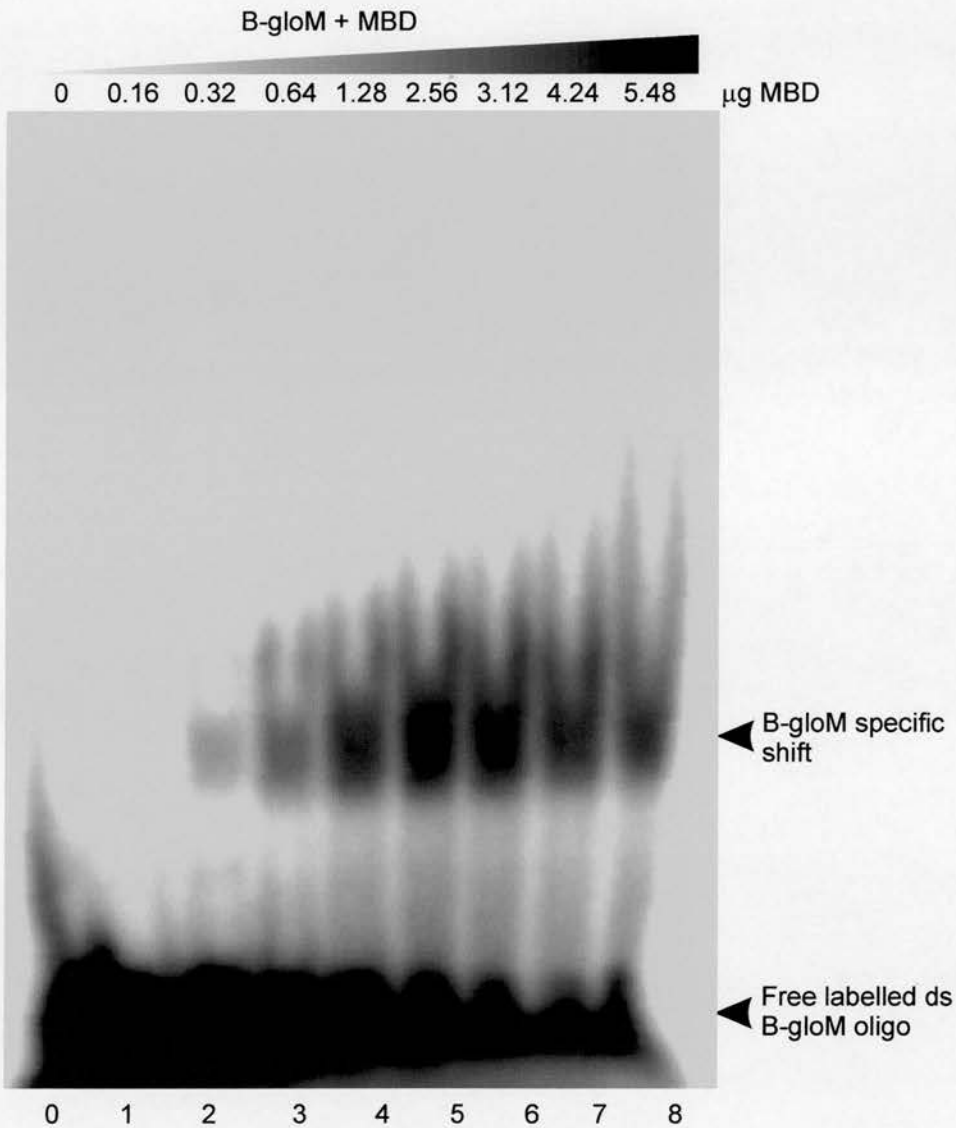


Figure 5.9 Bandshift analysis of the purified MBD protein with the B-gloM oligomer. Double-stranded B-gloM labelled DNA oligomer was incubated with increasing concentrations of the purified MBD protein. The free DNA and bound DNA fractions are indicated in the text by filled arrowheads. Lane 0 is a control lane lacking the MBD protein. Samples were run on a 5 % polyacrylamide gel.

A

```
hMBD1       LDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFKQGIL
xMBD1       EDWPLLGPGWKRRNVVRKSGATCGHSDTYYRSPAGKKIRSRIELAKYLGSAVDLSFFDFRNGVI
             * * ******** * ******** ***** ** *  ***  **  *** * **  ***  *
```

B

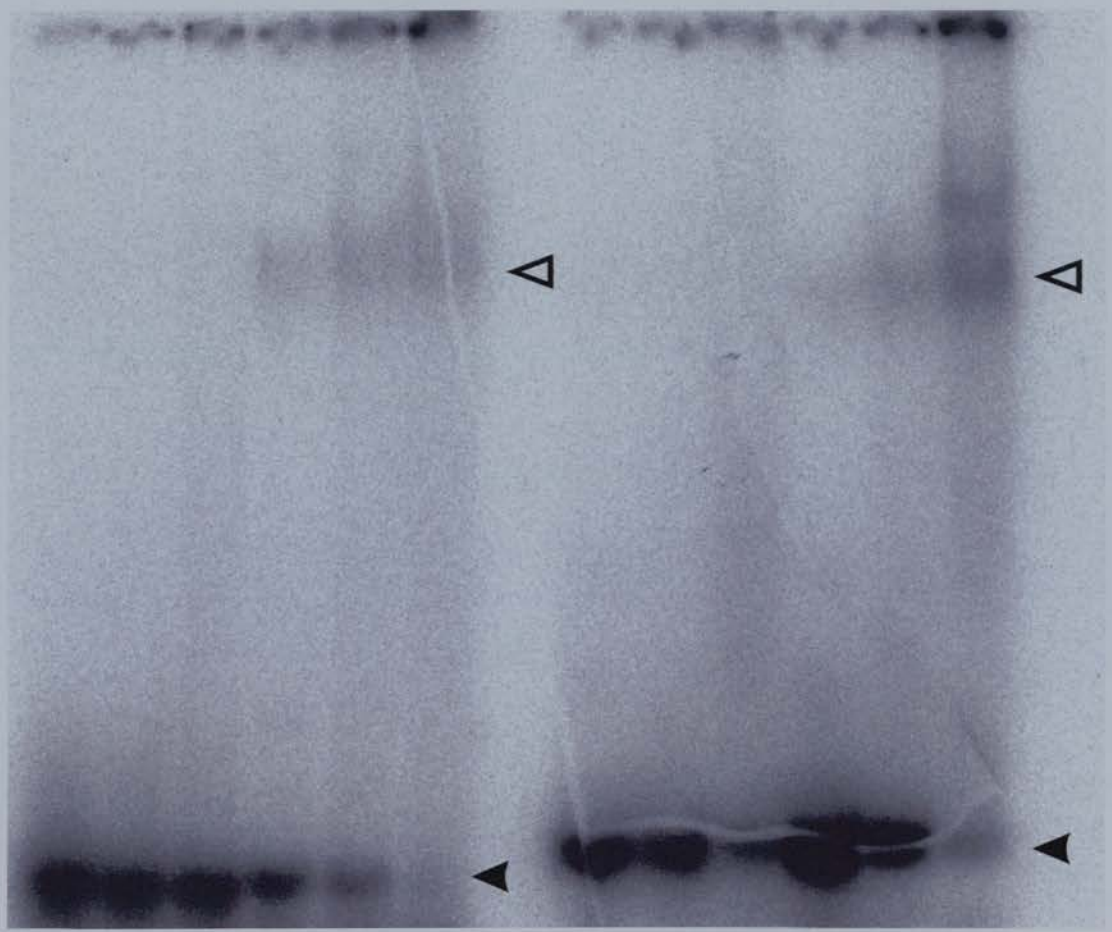


Figure 5.10 Bandshift analysis of the purified xMBD1 protein with the A-dnaU and A-dnaM oligomers. ***A*** amino acid alignment showing the degree of similarity between the human MBD1 and the xMBD1 proteins. ***B*** The purified xMBD1 protein provided by Dr. I. Stancheva, was incubated at increasing concentrations with a constant amount of either double stranded labelled A-dnaU and A-dnaM oligomers. Filled arrowheads indicate free DNA. Open arrowheads indicate a specific xMBD1 bandshift with both the A-dnaU oligomer and the A-dnaM oligomer. Samples were run on a 5 % polyacrylamide gel.

methylated lane, or perhaps the xMBD1 activity is low in this purification. It is also very possible that this result reflects the fact that xMBD1 can interact with A-form DNA. The number of methylated CpGs may be too few, or not in the correct sequence context to show enhanced binding with the CpG methylated A-dnaM. There is also no competitor DNA used in this study, but nevertheless some interaction is observed.

5.1.4 Footprinting analysis of the β^A-globin gene promoter region with purified MBD protein.

Having analysed the CpG triplet of the chicken β^A-globin gene promoter region in detail and performed bandshift analysis as described above, I decided to attempt DNaseI footprinting analysis on the β^A-globin gene promoter region using the MBD domain of MeCP2 that I had purified. The aim is to bind the protein to the DNA sequence and digest away the sequence surrounding the protein binding site. Initial tests estimated a concentration of DNaseI that would cleave approximately 50 % of the DNA sequence (data not shown). DNA fragments were prepared as described for Benzonase analysis of the β^A-globin gene promoter region. Fragments were labelled at the *Hinf*I end of *Sss*I methylated DNA. The DNA was purified and aliquoted into Bandshift buffer with increasing concentrations of MBD protein on ice. An equal amount of DNaseI was added to each sample and the digestion reaction carried out at room temperature. Digested fragments were purified and run on an 8 % denaturing PAGE/urea gel. The gel was dried and exposed to a Fugifilm phosphorImaging screen. After two days the screen was phosphorImaged. The resultant image is shown in figure 5.11. As can be seen from this image, there appears to be some binding of the MBD to DNA at the region of the CpG triplet. There is also another footprint downstream of the CpG triplet region. This may correspond to a single methylated CpG site located 45 bp downstream of the CpG triplet sequence. It therefore appears that the MBD of MeCP2 is capable of binding specifically to both singly methylated sites and the methylated CpG triplet region *in vitro*.

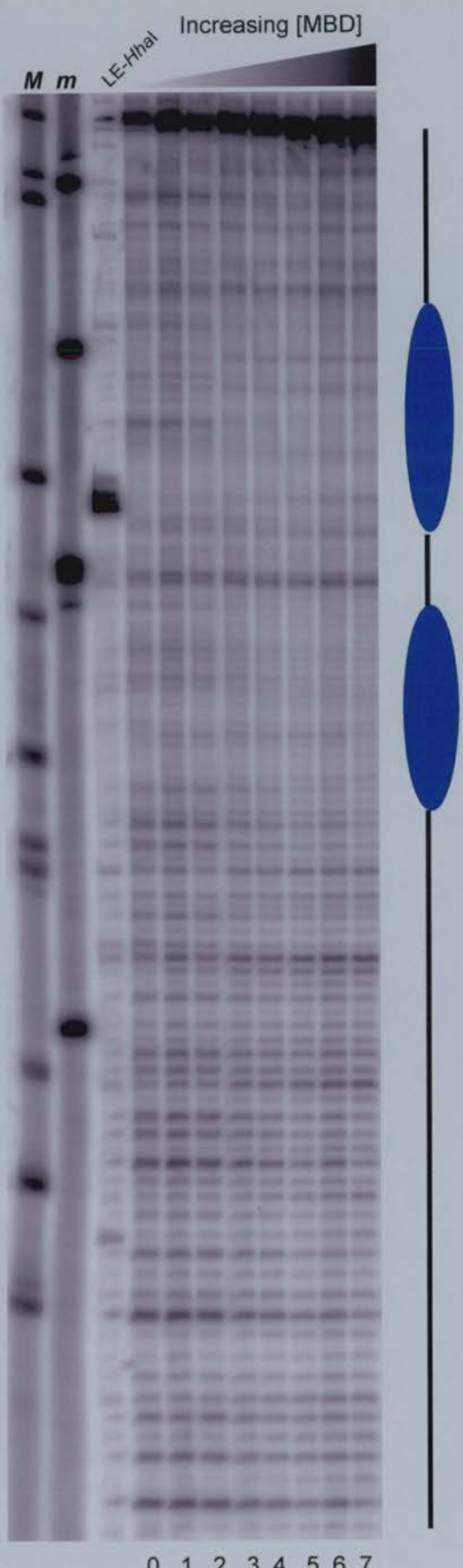


Figure 5.11 DNaseI footprint analysis of methylated β^A-globin LE fragments with increasing amounts of MBD protein and a constant amount of DNaseI. M, pBR322-*Bsu*RI digest marker; m, pUC19-*Msp*I digest marker; LE-*Hha*I, LE fragment digested with *Hha*I. Elipsoids show the protection in the sequence due to MBD binding. DNaseI concentrations were the same in each digested lane as was the quantity of radiolabelled DNA. MBD concentrations were increased across the gel.

5.1.5 Summary

These Bandshift experiments are initial attempts to unravel the specificities of binding of the methyl-binding proteins. Here I have demonstrated a distinct preference for methyl-CpGs by the MBD protein. It is irrelevant whether there are one or three methylated sites present in the bound sequence. It binds equally well to both the double-stranded A-dnaM and B-gloM oligomers. It appears that the MBD can specifically recognise the methyl-CpGs themselves and the effect they produce at a local structural level, ie: at the level of the dinucleotide pair. This is consistent with the known structure of the MBD and the determination of how it interacts with DNA.

It also appears that the MBD can bind the unmethylated A-form A-dnaU oligo to form a discrete complex, albeit at high concentrations of MBD only. This reactivity is reduced significantly in the presence of non-specific competitor. The MBD also shows some reactivity with the B-dnaU sequence, but this interaction is weak and smeared across the gel.

The xMBD1 protein is most similar to human MBD1. When used in the bandshift assay it appears to react poorly with the methylated A-gloM fragment when compared to the unmethylated fragment. As stated, this may be for several reasons but could provide an example of an MBD protein that can recognise methylated DNA when it is presented as an A-form structure.

Footprinting analysis of the β^A-globin gene promoter sequence (see Chapter 4 for details) clearly shows that the MBD is capable of locating to CpG methylated DNA *in vitro* over the triplet region and at high concentrations, has a weak affinity for unmethylated DNA. It is possible that *in vivo* the MeCP2 protein can locate to this sequence, although it is currently not known what methyl binding proteins operate at the promoter region *in vivo.*

A series of questions were posed in section 5.1.3 on how the MBD domain recognised methylated CpGs. To reiterate: It is known that MeCP2 can recognise a single methyl CpG site and that the MBD is responsible for this interaction (Meehan *et al.,* 1992, Nan *et al.,* 1993). Is this dependent on the sequence surrounding the CpG step? Is it the A-form structure of CpG methylated DNA that the MBD recognises and binds to? Can the MBD therefore recognise unmethylated sequences

that potentiate A-form DNA? To summarise and answer these questions, it is (a) clear that the MBD of MeCP2 alone can recognise and bind to methylated DNA, (b) clear that the MBD of MeCP2 recognises the CpG alone and does not seem concerned about changes in DNA structure surrounding a meCpG step, (c) apparent that A-forming sequences alone are not sufficient to appreciably compete for the binding of the MBD protein or the xMBD1 protein, and (d) not clear whether other MBD proteins that can only bind to multiple meCpGs (the MeCP1 complex), do so because they recognise methylated A-form DNA.To thoroughly prove that A-DNA has a role in the determination of where and when methyl-binding proteins can bind to their target DNA, a great deal more experiments are required. Presented in this thesis is proof of the existence of A-form structures in methylated DNA. It is clear that in some cases the formation of A-DNA can alter nucleosome positioning. From other studies it is also clear that different MBDs have different methylated sequence preferences. It may be the case that the effect of where and when methyl-binding proteins bind to their cognate sequence, is dependent on the response of nucleosomes to an altered DNA structure. Bandshift analysis can provide us with information on the methylation density that different methyl-binding proteins can bind to. To investigate this further, a variety of other purified MeCPs would need to be tested thoroughly for their binding preferences. A specific set of immediate questions need to be addressed. Firstly, do any of the other methyl-binding proteins, (hMBD1, xMBD3, MBD4 and Kaiso) show distinct differences in their binding capacities? Do any of these proteins show a distinct preference for methylated A-form DNA? If not, it is possible that an undetermined component of the MeCP1 complex of proteins (see fig. 1.7), is responsible for the direction of where and when different methyl-binding proteins bind. Another possibility is the highly flexible, yet variable loop structure present in the MBD1 and MeCP2 protein structure. The loop changes conformation on binding to different methylated cytosine residues. This may determine what further interactions are possible in terms of higher order chromatin. A combination of competitive bandshifts with a greater range of both DNA and RNA sequences would need to be performed to reveal individual MBD protein specificities. It may be that competition with dsDNA A-forming sequences is only observed in the presence of the MeCP1 complex (see figure 1.7). Bandshift

experiments with the purified MeCP1 complex from Xenopus oocyte extracts could shed some light on this distinct possibility. A further great advantage in understanding the mechanism of MBD protein binding, will be the inevitable elucidation of the NMR structures of MBD interactions with methylated DNA. Indeed, we are still not sure how many methyl-binding proteins currently exist. Until now a selection have been identified that contain an MBD. Kaiso is one example of a methyl-binding protein that does not contain an MBD domain. There may be others in existence that use these extraordinary CpG methylation-induced transitions in DNA structure to their regulatory advantage.

Chapter 6: Discussion

Note: I have included a Summary at the end of each chapter where I have discussed the results in each chapter. Therefore I have used this chapter to both summarise the data as a whole and speculate on the potential implications of these findings.

The subject of this thesis has been an investigation into the structure of CpG methylated DNA. As described in Chapter 1, there is a lot of evidence to support the idea that DNA methylation can alter DNA secondary structure. There is also substantial evidence to support the idea that the cytosine residue is crucial in determining the equilibrium between B-DNA, Z-DNA and A-DNA (Foloppe *et al,* 1999). Only recently was a naturally occurring nuclease from *Serratia marcescens*, Benzonase, discovered with the ability to cleave DNA with a distinct preference for A-form DNA. This enzyme had never been applied to CpG methylated DNA before. Here it is demonstrated for the first time, that CpG methylated DNA can potentiate the formation of A-DNA *in vitro,* using an assay based on the specificity of the DNA/RNA nuclease, Benzonase.

6.1 The Benzonase Enzyme: History and specificity for A-form DNA.

The *Serratia marcescens* extracellular endonuclease is a 266 amino acid protein that is proteolytically processed upon excretion to give an active protein of 245 amino acids (Ball *et al.,* 1987). It is commercially available as Benzonase and is used in industry for the downstream processing of pharmaceuticals. The structure has been solved by crystallographic analysis (Miller *et al.,* 1994) and shown to be a dimer of identical subunits (Friedhoff *et al.,* 1994), although mutational analysis shows that dimerisation is not necessary for activity (Franke *et al.,* 1998). It is an extremely stable enzyme with a temperature optimum between 37 and 44°C and requires a divalent cation such as Mg^{2+} for activity. The commercial product was obtained from Merck as a 30 kDa protein which has been purified to purity grade I: 99 %, with no

detectable protease activity present. It is supplied at 25 U/µl and is incredibly potent. 1 U can degrade 37 µg of DNA in 30 minutes (Merck, technical information).

Benzonase has a broad substrate range. It can cleave both single and double stranded DNA and RNA, producing 5'-phosphorylated mono-, di-, tri and tetranucleotides (Nestle & Roberts, 1969a, 1969b). Meiss *et al.,* (1995) have investigated the specificity of Benzonase on different types of DNA and RNA and found that although it will cleave all nucleotide sequences, it has preferred cleavage sites. Other non-specific nucleases show similar specificities. For example, DNaseI has sequence preferences that are explained by both local and global variations in DNA structure (Drew & Travers, 1984; Fox, 1986) as well as DNA flexibility (Hogan *et al.,* 1989) and DNA bending (Brukner *et al.,* 1990). For DNaseI, nuclease-olionucleotide complex structures have been crystallised and therefore it is possible to rationalise sequence preferences (Weston *et al.,* 1992). Only the three dimensional structure has been solved for Benzonase alone, without oligonucleotide binding; by Miller *et al.,* (1994), refined by Lunin *et al.,* (1997) and more recently to a resolution of 1.1 Å by Shlyapnikov *et al.,* 2000). Meiss *et al.,* (1995) have compared the specificity of Benzonase and DNaseI for a series of oligonucleotides, none of which contained CpG methylated DNA. They observed marked differences in preferred cleavage sites. The most notable difference observed by Meiss *et al.,* (1995), was that Benzonase has preferential cleavage of d(G).d(C) tracts, which are not attacked preferentially by DNaseI. Both have an inability to cleave dsDNA at d(A).d(T)-tracts, a sequence that has a small minor groove width. The minor groove can vary in width quite substantially. For example, the sequence d(CGATTAATCG) has a narrow minor groove width measuring 2.6 Å, whereas the sequence d(CCGGCGCCGG) has a minor groove width of 8.9 Å (Heinemann *et al.,* 1994). Benzonase, like DNaseI attacks DNA at the minor groove. This is speculated on by Meiss *et al.,* but it is also seen in thesis, that in sequences where DNA is preferably cleaved by DNaseI, there is often no overlap with Benzonase (chapter 3). When the two enzymes do converge with the same digestion pattern, it is at regions known to have a much larger minor groove than average, that fall under the classification of A-DNA. It appears that the minor groove must be widened more for Benzonase than for DNaseI, for it to show enhanced cleavage (Meiss *et al.,* 1995, and this thesis).

In line with this, further investigations by Meiss *et al.*(1999), have confirmed that Benzonase does indeed prefer A-form nucleic acids as substrates. Steady-state kinetic analysis of the cleavage of a selection of olgonucleotides, in conjunction with structural data from circular dichroism spectroscopic analysis, suggests that oligonucleotides adopting an A-conformation are preferred over those that potentiate a B-conformation. Both single-stranded and double-stranded substrates of both A-DNA and B-DNA were used in this study. The substrate showing the best cleavage kinetics was a dsRNA oligomer of the sequence called r(r)= r(GCAAAAAAGCG), which is known to form an A-form structure. A feature of the A-conformation is a C3'-endo sugar pucker conformation and a tilt of the bases relative to the helix (Olson, 1975). These features are found in this oligomer and in others that also reacted well with Benzonase, but were intermediate in their establishment of the A-conformation. This aspect of the reactivity of Benzonase has been a vital understanding in terms of this thesis. If Benzonase can recognise the C3'-endo pucker then this provides us with a fairly specific A-DNA nuclease. This work also revealed that Benzonase prefers double-stranded substrates over single-stranded, rendering it a useful tool for the analysis of methylation induced A-form DNA, as it can ignore any unusual secondary structure abnormalities or regions of single-strandedness in the sequence (Meiss *et al.,* 1999). Further analysis to confirm this was performed on a range of dsDNA oligomers that had the propensity to form either A-form or B-form DNA. In all cases, Benzonase reacts with the phosphodiester bonds of the A-conformation oligomers much more readily (Meiss *et al.,* 1995, 1999). Of special note is that although the authors declare there is a cleavage preferance for G+C rich regions, this does not reflect a base preference. The oligo d(G).d(C) substrate is cleaved preferentially and has a greater tendancy than oligo d(GC) to be A-form (Heinemann *et al.,* 1987, 1994; Meiss *et al.,* 1999). Cleavage in alternating d(CG) sequences that adopt the B-conformation in physiological conditions is low (Butkus *et al.,* 1987).

Interestingly, in the analysis of single stranded substrates it was found that the sites of hydrolytic attack are altered when the oligonucleotide is rendered double-stranded. The variations indicate that Benzonase is sensitive to local structural variations of the phosphodiester backbone of nucleic acids. Further to this,

mutational analysis of conserved amino acids at the Benzonase binding site do not affect its activity a great deal. The results of several amino acid substitutions at these conserved sites confirms that substrate recognition covers the whole substrate binding site and is not the result of a particular residue at the catalytic centre (Meiss *et al.,* 1999). Based on these many observations by Meiss *et al.,* (1995, 1999), I was convinced of its ability to recognise and cleave A-form DNA sequences with a great deal of specificity. I decided to apply this knowledge of Benzonase reactivity to CpG methylated DNA to see if it too would show enhanced cleavage, in comparison to unmethylated DNA. This would indicate regions of potential A-form DNA, not only providing a new type of assay for the analysis of methylated DNA, but also underlining a potential role for A-form DNA in gene regulation.

6.2 CpG Methylation Potentiates the formation of A-form DNA.

The hypothesis was that CpG methylation could potentiate an A-form structure. This could make CpG methylated DNA a preferential target for Benzonase digestion. Heavily methylated DNA might therefore be digested faster than non-methylated. A pilot experiment used Benzonase to digest unmethylated and *Sss*I methylated pBR322 plasmid (figure 3.2). There did not appear to be any significant increase in the overall rate of digestion by Benzonase. I concluded that although the enzyme may be expected to digest methylated DNA faster than unmethylated DNA, this effect would not be observed unless the plasmid was incredibly CG rich. There were enough unmethylated sites in pBR322 to keep the overall digestion rates constant. This did not exclude the possibility that at the level of the methylated base pair some local effect could be observed. This aspect was pursued as described below.

I decided to use Benzonase to develop an assay similar to that used by Meiss *et al.,* 1995, 1999. To establish that the method I was developing would be effective in detecting regions of A-form DNA, a variety of tester fragments were used, illustrating the ability of this assay to detect sequences known to potentiate A-DNA. The first of these is a fragment of the pBR322 plasmid covering the ampicillin resistance gene (Amp^R). The region covered in my assay is larger that used by Mei *et al.,* (1988) who used a Ruthenium based chiral probe to detect A-DNA. The region that does overlap with their analysis, shows the same digestion hotspots in the

unmethylated fragment using Benzonase as Mei *et al.,* (1988) observed using their Ruthenium probe, confirming the usefulness of Benzonase for A-DNA analysis. When this sequence is CpG methylated, the pattern of specificity of Benzonase changes (figure 3.5 *A*). Although there are several methylatable CpGs in this fragment, there is only one region that shows enhanced cleavage with Benzonase. This corresponds to a region covering a *Hha*I recognition site and covers the sequence 5'..GGTTCCGCGCAC..3'. There are two methylatable CG sites adjacent to each other here that are responsible for the altered cleavage at this site. This is the first example of the detection of methylation induced A-form DNA using this assay. In a later additional experiment I decided to analyse the reverse strand of this fragment. There is a structural change in the sequence at the expected location although the marker lane nor the separation of bands in the gel, is as good as I had hoped. It is therefore included in the Appendix, Chapter 7 as figure 7.1.

*Sss*I methylase methylates every CG dinucleotide pair in the pBR322 fragment, and was confirmed to do so by a methylation protection assay (data not shown). DNaseI shows enhanced cleavage at methyl CpGs as determined by Fox, (1986). I performed DNaseI analysis on this fragment to confirm that there were structural alterations present in the DNA sequence due to methylation, that were not being picked up by Benzonase. DNaseI analysis of the pBR322 gene fragment detects that CpG methylation ***does*** induce DNA structural changes at each methylated cytosine, including the site that shows enhanced cleavage with Benzonase (figure 3.6 *A*). This experiment confirms that the Benzonase enzyme requires a larger change in structure than DNaseI requires, for it to react preferentially. Both DNaseI and Benzonase are sensitive to groove width as described above. It appears that Benzonase requires a larger widening of the minor groove of DNA to show preferential cleavage at methylated sites, a feature that is consistent with A-form DNA.

Another tester fragment was then subjected to Benzonase analysis. An 167 bp insert (the 'Pingoud' fragment) into pGEM9Zf(-) was designed and kindly provided by Dr. Pingoud. This fragment had been used by Meiss *et al.,* (1999) to demonstrate the sequence specificity of the *Serratia* nuclease and was part of their analysis that concluded that the nuclease could recognise and cleave A-form DNA

preferentially. In my analysis I have proven that my Benzonase reaction conditions provide the same pattern of digestion as that shown by Meiss *et al.,* (1999) for this fragment in the unmethylated state. However, when I methylate this fragment the pattern of digestion changes (figure 3.8). Now some sites show enhanced cleavage where they were largely unreactive in the unmethylated fragment. The major difference seen is at a $d(GC)_n$ string that it is largely unreactive in the study by Meiss *et al.,* but is highly reactive in my analysis when this sequence is CpG methylated. This confirms again that Benzonase is indeed a useful tool for the analysis of methyl CpG induced DNA structure. This time it is capable of recognising A-form structures in a much longer sequence context. Long alternating $d(GC)_n$ are commonly found at CpG islands at gene promoters. These are usually unmethylated at gene promoters, apart from at the inactive X-linked gene loci and some imprinted genes. Normally unmethylated CpG islands can become aberrantly methylated rendering the gene inactive. CpG island methylation can sometimes lead to the progression of cancer, especially if the silenced gene is involved in the processes of apoptosis, DNA repair or cell-cycle control mechanisms (Jones & Baylin, 2002). It is possible that A-DNA may form at aberrantly methylated CpG islands *in vivo,* which is probably then recognised by the methylation machinery.

Having established that CpG methylation can potentiate A-form DNA, I then tried to refine the conditions of the experiments by using Urea. Meiss *et al.,* (1999) had reported that Urea increases specificity of the enzyme. I attempted this with both the pBR322 fragment and the Pingoud fragment. However I did not see any enhanced specificity in my assays with Urea. (figure 3.9 and 3.10). Interestingly for the commercial Benzonase documentation from Merck, they report that any increased specificity seen by the use of Urea is a pseudo-effect and reflects the digestion of DNA in these conditions. Perhaps the timescale used in my analyses was not sufficient to show digestion by Urea, but sufficient to show digestion due to Benzonase. In all methylated fragments however, Benzonase did show heightened reactivity at the same sequences in the presence of Urea, confirming that the changes seen were not due to the reaction conditions, and were not altered in the presence of Urea. This may indicate that the methylation induced structural transitions recognised by Benzonase are indeed very stable. The reactivity of Benzonase at

various methylated sites in DNA is summarised in table 6.1 below, along with a list of DNA sites recognised by DNaseI analysis, that are not recognised by Benzonase.

(+) Benzonase Cleavage (+) DNaseI	Data Source
CCGCGC	Figure 3.5 *A*
CGCGCGCG	Figure 3.10
GGGCGCCC	Figure 3.10
GCGCGC	Figure 3.13
GCGCGG	Figure 3.13
GGCGC	Figure 3.13
GCGCGG	Figure 3.13
GCGCGC	Figure 4.3
(+) DNaseI (-) Benzonase Cleavage	**Data Source**
CG	Figure 3.6 *A*
GGCG	Figure 3.6 *A*
CCCCG	Figure 3.6 *A*
GCGG	Figure 3.6 *A*

Table 6.1: Benzonase and DNaseI digestion: Preferred cleavage sites at CpG methylated DNA sites. The above table indicates sites shown to react with Benzonase from Results, Chapter 3 and one example from Results, chapter 4. **(+)** indicates enhanced cleavage with either Benzonase or DNaseI. **(-)** indicates sequences that do not show enhanced cleavage with Benzonase, but do show enhanced cleavage with DNaseI. The data source refers to the various gels where reactivity has been observed. All of the sequences were *Sss*I methylated, therefore although not indicated in the sequences above, it is assumed that all of the above sequences are methylated at CpG dinucleotides.

Having established the technique and some of the sequence preferences of Benzonase for methylated DNA, I decided to focus on a eukaryotic gene sequence, where DNA methylation is known to have an effect on gene expression. The α_2-globin gene was chosen for several reasons. It is highly CG rich and possesses a CpG island, which is usually unmethylated, but can become methylated in some cell lines, where they can subsequently bind MeCP1 (Boyes & Bird, 1991, Cuadrado *et al.*, 2001).

The developmental regulation of globin gene expression is intensely studied because of their importance in many common inherited diseases, in particular the thalassemias. A high degree of conservation of these developmental control mechanisms has been observed from human to avian species. In vertebrates, the α– and β– globin genes are expressed sequentially and exclusively in erythroid cells during defined stages of development. The human α–like gene family is comprised of a gene cluster in a 26 kb region on chromosome 16 in the order (5'-ζ(embryonic)-α_2 (adult)-α_1 (adult)-θ_1(foetal/adult)-3') and β-like gene family on chromosome 11 (5'-ε(embryonic)-Gγ(foetal)-Aγ(foetal)-δ(adult)-β(adult)-3'). The avian β-type globin cluster (5'-ρ(embryonic)-βH(adult)-β(adult)-ε(embryonic)-3') is arranged differently with the two embryonic genes flanking two adult genes. Both α- and β-globin gene sequences regulate the timing of gene expression *via* locus control regions (LCRs), erythroid specific enhancers, located kilobases upstream of each gene cluster (review: Weatherall, 2001). Figure 6.1 illustrates the organisation of the human α- and β-globin gene families. The chicken β-globin locus is discussed later, but like the human globin genes sweeping changes in gene expression patterns are also observed. This figure is used to illustrate the dramatic changes in expression that occur throughout early development. CpG methylation could provide part of the explanation as to how these genes are expressed and shut down at various stages of development. At present it is not understood how CpG methylation may be involved in developmental changes of gene expression at the globin loci.

There are many *Hha*I sites in the α_2-globin promoter region, which had already been shown to be very responsive to Benzonase in the studies thus far. Benzonase analysis of a region of the α_2-globin promoter region shows the most reactivity seen with Benzonase to this point (figure 3.13). There are many sites that show an altered structure due to CpG methylation. These mainly consist of *Hha*I recognition sites, although there are two overlapping *Hpa*II sites that seem to show some reactivity with Benzonase. At this region between 80 and 89 bp, marked in figure 3.13, it is possible that *Hpa*II recognition sites do contribute to a change in structure when methylated. It is also possible that because there is some alternation of sequence, and at least one partial *Hha*I site at this region, that it is not the *Hpa*II sites *per se* that are being recognised, but an overall methylated CpG richness. To

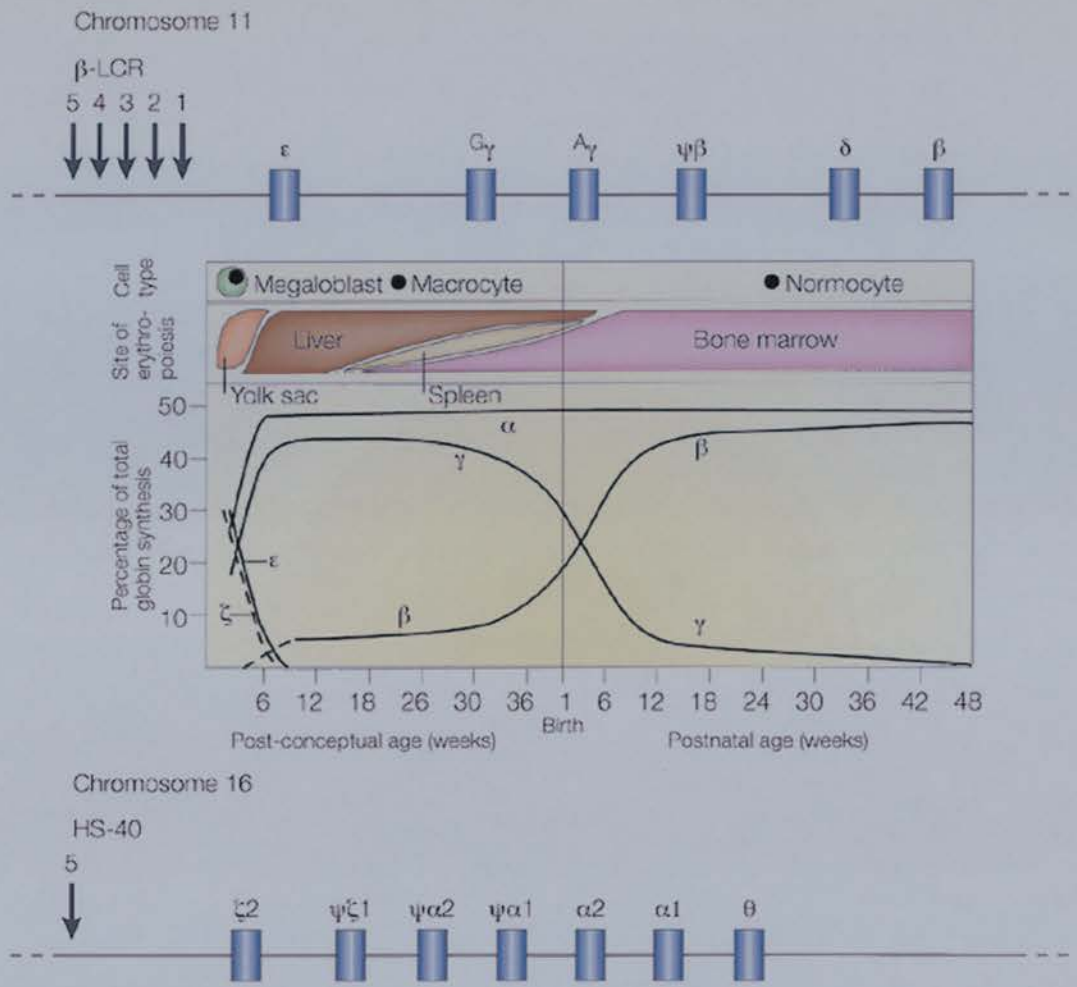


Figure 6.1 The chromosomal organisation of the human globin genes.
This diagram illustrates the organisation of the globin gene families. Vertical arrows indicate the location of DNaseI hypersensitive sites that are thought to be involved in globin gene regulation. The β-globin family is clustered on chromosome 11 and the α-globin family on chromosome 16. LCR, locus control region. HS-40, Hypersensitive site 40. The products of the $^{G}\gamma$- and $^{A}\gamma$- genes are g-chains with either glycine ($^{G}\gamma$) or alanine ($^{A}\gamma$) at position 136. The central diagram summarises information on the levels of expression of the α-,β - and γ-globin genes through early development. This information is coupled to both the tissue of expression in the body and the cell-type where expression is found. A megalocyte is a large red-cell precursor, a macrocyte is a large red cell and a normocyte is a normal-sized red cell. Of note is the major shift in expression of the globin genes throughout development. At about 12 weeks, post conception, the expression of the β-globin gene gradually rises until it reaches a maximum circa 36 weeks of post-natal development. The reverse pattern is seen for γ-globin gene expression. It reaches a minimum at 36 weeks post-natal development. The α-globin gene maintains a steady state of expression from 12 weeks post-conceptual development. The most dramatic changes in gene expression occur in early post-conceptional development. Embryonic ε-globin rises and falls dramatically before the 12 week stage, at which point the α-, β- and γ-genes come into play. It is these sweeping changes in gene expression that the interplay of CpG methylation and histone/chromatin modifications are proposed to control. (Adapted from Weatherall *et al.*, 2001).

confirm that these changes were due to heightened reactivity at alternating $(GC)_n$ regions, the experiment was repeated, but this time using *Hpa*II methylase to methylate the fragment. The *Hpa*II recognition sequence is not an alternating sequence. In this experiment I observed that there were now no observable highly reactive sites in the methylated sequence as compared to the unmethylated sequence (figure 3.14). This implies that Benzonase is only highly reactive with some of the methylated sites and not others. Sequences that show alternation of GC residues seem to react well with the enzyme. These sequences in their methylated state are also the most consistent A-forming structures, as previously discussed. For the relevance of these results to gene regulation I have investigated the literature for relevant information on the regulation of globin gene expression, as discussed below.

A map encompassing 300 kb in and around the human alpha-globin cluster also shows that a 26 kb cluster is also embedded within a region of unusually high CG content (Fischel-Ghodsian *et al.*, 1987). Like other tissue-specific CpG island genes, both α_1 and α_2 genes are completely unmethylated in non-erythroid and erythroid tissues tested including sperm, but are only expressed in erythroid tissues (Bird *et al.*, 1987; Antequera *al.*, 1989; Vyas *et al.*, 1992). One major difference in α-globin gene regulation is that both the human and rabbit alpha-globins can be expressed in transfected erythroid and non-erythroid cells in the absence of an enhancer. This enhancer independent expression of the alpha gene requires extensive sequences not only from the 5' flanking sequences but also from the intragenic region. Whether this is due to an intragenic enhancer with binding sites for gene specific proteins or due to an extension of the promoter into the intragenic region was addressed by Shewchuk *et al.*, (1997). They concluded that both possibilities above were not likely by their analysis, but that the CpG island extended into the intergenic region and the effect seen was a general one; *ie:* When the size of the CpG island is increased, the level of gene expression increases concomitantly. They also show that the CpG island provides a chromatin environment that promotes transcription.

Apart from the aforementioned evidence on CpG island methylation causing gene inactivation in cell lines, accumulating evidence for natural expressing cells also exists. This identifies the interweaving of both DNA methylation and

chromatin structure as of major importance in the regulation of both tissue-specific and developmental stage specific expression of the α-globin gene cluster. Throughout the α-globin gene family there are many DNaseI hypersensitive (HS) sites, which correlate with changes in chromatin structure at different developmental stages (Yagi *et al.,* 1986). These changes also correspond with sequential changes in the DNA methylation status 5' to the coding region of the α-globin genes in the switching from embryonic > foetal > adult haemoglobin gene expression (Haigh *et al.*, 1982; Michalowsky *et al.*, 1989). A major LCR is located 40 kbp upstream of the human embryonic ζ-globin gene. It too is required for the expression of the α_2–globin gene and most of the chromosomal deletions that cause alpha-thalassemia either remove the alpha-globin genes and/or the LCR (HS-40) (Barbour *et al.*, 2000). *In vivo* footprinting of this region identified binding sites for GATA-1, AP-1/NF E2 and unknown CACC / GGTGG binding proteins all of which are required for correct LCR function (Strauss *et al.*, 1992). In terms of protein binding, Rein *et al.*, 1995, have demonstrated that NF-1 also binds, and is partly responsible for alpha globin repression, but concluded that there may also be different forms of NF 1 repressing alpha-globin expression (Rein *et al.*, 1995). Additionally, in the 5'-flanking and internal regions of the rabbit α-globin gene, which constitutes the CpG island and confers enhancer-independent transcription, contains binding sites for transcription factors alpha-IRP (-53 to –44), CP-1 (-73 to –69) and Sp1 (-95 to –90) relative to the CAP site. Sp1 also binds to sequences in intron 1 and unidentified proteins also bind to CCAC and C5 motifs in the 3' UTR. Exon 1 also contains a binding site for the nuclear protein YY1. The human alpha-globin gene also contains multiple copies of the Sp1 binding sequence. Interestingly deletion of the alpha-globin 5' flanking region in transient transfection assays, has no effect on α-globin gene expression. However, when the deleted construct is introduced into chromatin, loss of the 5' region decreases expression by 90 %. Footprint analysis and functional assays show that the ability of the 5' flanking region to increase α-globin gene activity is mediated in a chromatin context by its multiple binding sites for Sp1. Sp1 has a recognition site consisting of alternating d(CG) residues and does not lose its ability to bind a recognition site when it is CpG methylated *in vitro*, but perhaps interaction

with methyl-binding proteins may influence transcriptional regulation at these sites (Pondel *et al.*, 1995).

Finally, a recent study underlies the role that CpG methylation and associated proteins have in globin gene regulation, methylation of a 248 bp region proximal to the avian embryonic ρ-globin gene has been found to correlate inversely with stage specific expression in avian erythroid cells. Transient transfection of the CpG methylated 248 bp region into primary erythroid cells, in the absence of promoter methylation and in the context of nucleosomes, forms a methyl-specific complex containing MBD-2 and HDAC-1 proteins (Singal *et al.*, 2002). Similar mechanisms may be involved in the sequential expression of the α-globin genes and others throughout development.

My attention then shifted to a gene that could provide information pertinent to both DNA and chromatin structure. As discussed in Chapter 4, this involved analysis of the chicken β^A-globin gene. This gene had previously been analysed by Dr. C. Davey. It contains a CpG triplet sequence in the promoter region that causes alterations in DNA structure when CpG methylated in addition to nucleosome displacement (Davey *et al.*, 1997). I began a collaboration with Dr. Davey on the nature of this structural change. Dr. Davey analysed the region for the formation of Z-DNA but found that it could only form on a supercoiled template and not on linear DNA. Nucleosome displacement occurs on linear DNA and is therefore not due to Z-DNA formation (pers. comm.). I analysed the same gene region with Benzonase digestion analysis for the presence of A-DNA (figure 4.3). The results clearly show a methylation dependent structural change at the position of the CpG triplet. In addition, Benzonase analysis of the reverse strand of this sequence also showed a methylation dependent change, (figure 4.4) indicating the change in structure was a consistent feature of both strands and less likely to be some kind of secondary structure alteration located on one side of the helix. From my analysis it appears that this region of the chicken β^A-globin gene can form A-DNA, which coincides with the region responsible for methylation induced nucleosome displacement.

A series of point mutations and insertion mutations of the CpG triplet were analysed by Benzonase digestion to assess their A-forming potential, and to ascertain whether this coincided with nucleosome mobility or not. Dr. Davey performed

DNaseI analysis and nucleosome mapping (monomer extension) on these mutants. The details of these are discussed in Chapter 4. A summary of the data and the sequences of the wild-type and mutant fragments are shown in table 6.2. The pattern that emerged from Benzonase analysis of the mutated CpG triplet sequences was one of a graded response, dependent on how many methylated CpGs are present, whether they lie in adjacent positions and whether there is alternation of base sequence around the methylated sites. All of the methylated mutants retained some ability to show enhanced reactivity with Benzonase but sometimes provided slightly different cleavage patterns or different band intensities than the wild type. For example, the

Name	Sequence	Benzonase	DNaseI	Nucleosome Displacement
LE	CGCGCGCTGTGCT	*****	*****	*****
m1	GCCGCG	****	****	**
m2	CGGCCG	***	***	***
m3	CGCGGC	****	***	****
m23	CGCCCG	***	****	*
m25	CGGGCG	***	***	***
m8a	CGGCCGCTGTGCG	***	***	***

Table 6.2 Tabulated information for Benzonase, DNaseI and Nucleosome displacement analyses. Benzonase analysis was performed as described in this thesis on the β^A-globin gene promoter and a selection of mutations in the CpG triplet. The star rating is as follows. Five stars, very strong reaction with Benzonase, DNaseI and ability to displace nucleosome 5A (see chapter 4). Four stars, very good; Three stars, good; Two stars, poor; One star, negligible. These ratings are based on unmethylated and methylated fragments in a side by side analysis. They are a guide only and are not quantitatively analysed.

m1 mutant has two adjacent meCpGs, the first of which is not a good A-forming sequence as it begins with a pyrimidine (see table 1.2). This is reflected in the Benzonase pattern (figure 4.7). There are at least two main hypersensitive sites in the methylated m1 fragment that correspond with the methylated LE fragment. One of these bands is less intense than it appears in the methylated LE fragment. In the methylated LE fragment the strong central hypersensitive band as indicated in figure 4.7 is most probably the centrally methylated CpG and is a strong hypersensitive site

because the methylated LE fragment contains three methylated CpGs and has a greater potential to form A-DNA. In agreement with the formation of A-DNA at the methylated LE CpG triplet, nucleosome 5A is displaced (figure 4.6 B). In accordance with some A-DNA formation, as judged by Benzonase on the m1 fragment, nucleosome 5A is again displaced but not as effectively. DNaseI analysis shows three strong hypersensitive sites over the CpG triplet for the methylated LE fragment (figure 4.5 *A*). For m1, there is one major DNase hypersensitive site and two minor ones, reflecting the fact that there are less methylatable sites in this fragment. It should be taken into account here that DNaseI activity on DNA will not detect A-form sequences, but it will detect methylated CpGs that cause structural alterations in DNA (Fox, 1986 and figures 3.5 *B* and 3.6 *B*). Therefore in many cases the DNaseI assay will detect sequences where the minor groove has widened, but does not account for the spreading effect that must occur as the helix changes from a B-form molecule to an A-form molecule. Benzonase is able to detect sequences that are partly A-form as reported by Meiss *et al.,* (1999), therefore for all of the mutants Benzonase may be detecting regions that are partly A-form and/or a result of nearby sequences that have adopted the A-conformation. As seen from previous Benzonase digests, (Chapter 3) two adjacent CpGs in an alternating sequence provides enough sequence to show enhanced cleavage at adjacent methyl CpGs. The CpG triplet mutations are over a very small region of 6 bp. This provides only a very small region to look for the formation of ***or*** lack of formation of A-DNA. The less likely a sequence is to form A-DNA from our structural knowledge of it, the less likely it is to react with Benzonase and demonstrate a pattern akin to the wild-type situation. However, as all of these sequences have some A-form potential, they are all expected to react with Benzonase to some extent, and they do. They are therefore all expected to move nucleosome positioning of 5A, and they do with varied success, apart from mutant 23.

Mutants 2, 25 and 8a all behave very similarly with respect to Benzonase analysis (figures 4.8, 4.9 and 4.10, respectively). They have a high degree of alternation of base sequence and are expected to form A-DNA from structural studies (Table 1.2). They all show enhanced cleavage in the Benzonase assay on the methylated strand, however the banding pattern is not as clear or as defined as we

see in the wild-type LE methylated fragment. The m2 and m25 fragments contain two methylatable CpGs. The m8a fragment contains the same two methylatable CpGs as the m2 mutant and also an additional CpG 6 bp downstream of the triplet. The additional CpG does not in my estimation change the behaviour of the m8a mutant in terms of Benzonase analysis or nucleosome positioning analysis; m8a behaves similarly to m2 in both respects. I believe the additional CpG downstream of the triplet sequence in m8a, is too far removed from the critical site of nucleosome occupation to have any major effect. The DNaseI data does acknowledge the presence of this extra methylated CpG however and but it does not change the DNaseI pattern at the mutated triplet sequence (figure 4.5 *A*). DNaseI analysis of all of these fragments shows hypersensitive sites at the same positions as the methylated LE fragment (figure 4.6 *B*), but they are not of the same intensity. As expected, nucleosome positioning analysis shows that all of these mutants are capable of shifting nucleosome 5A to some extent, as they all demonstrate some reactivity in the Benzonase assay.

The m3 mutant behaves the most similar to the wild-type situation. It contains two methylated CpGs with a high degree of alternation of base sequence and a purine start before the first methyl CpG, all features which help potentiate the formation of A-DNA (see table 1.2). In these terms it has the most similar structural characteristics to the wild-type situation and it is therefore expected that it should show the same kind of response in the Benzonase assay (figure 4.9). Benzonase analysis shows hypersensitive sites over the same three sites in the methylated m3 fragment as shown in the methylated LE fragment. Crucially there seems to be strong reactivity over the central band in this hypersensitive region, which probably represents the second methyl CpG in the mutated triplet and the corresponding central CpG in the methylated LE fragment. As there is a third less strong hypersensitive site in the benzonase analysis of the m3 mutant, this probably represents a spreading effect of A-form DNA potentiated by the first two methylated CpGs in the sequence. These first two methyl CpGs seem important for the positioning of nucleosomes. Only the LE and m3 mutants have methylation of the first two CpGs out of all of the sequences analysed. These two sites must be important for nucleosome positioning as when m3 is methylated at these two sites

nucleosome 5A is shifted with the same efficiency as in the wild-type situation (figure 4.6 *B*). DNaseI analysis of the m3 fragment reveals that it too has hypersensitive sites on the methylated fragment over the same three sites as the methylated LE fragment, one of these is strong and represents the first methyl CpG.

The last of the mutants analysed is m23. This provided the most revealing information from the mutants as it is not capable of shifting the position of nucleosome 5A (figure 4.6 *B*). All of the mutants show reactivity to Benzonase. The mutants LE and m3 are the most similar in their benzonase patterns. Both shift nucleosome 5A very efficiently. The mutants m2, m25 and m8a show the most similarity in reactivity to Benzonase and show similar efficiencies in moving nucleosome 5A. There are some differences in their ability to do this, but this can be explained by their A-form potential as defined by the degree of alternation of base sequence, whether the methylated CpG is preceeded by a purine (typically Guanine), and how many methylatable CpGs are present in this sequence. The mutant m23 however, does not effectively remove nucleosome 5A. There are two non-adjacent methylatable sites in this mutant. Benzonase analysis may show some reaction with but it is difficult to pick out the same three definable bands as observed in the methylated LE fragment. The pattern of unmethylated and methylated m23 fragments are very similar, and therefore methylation may be inducing little or no A-form structures.

The first methylated CpG in the m23 mutant has good A-form potential, but the second CpG has the lowest A-form potential of any of the sequences analysed. The second CpG is preceeded by two Cytosines, which detracts from the A-form potential of this sequence. Therefore the only real contribution to the formation of A-DNA may be in theory from the first CpG. In support of a central role for cytosine, it has been found that of all the deoxyribonucleosides, Cytosine has intrinsic conformational properties that may play a decisive role in the equilibrium between A, B and Z-forms of DNA (Foloppe *et al.,* 1999). Another important point to remember is that structures are now emerging that contain the properties of both A- and B-form DNA. The dodecamer d(CGCCCGCGGGCG) has been solved as a kinked A-DNA molecule with B-DNA features (Malinina *et al.,* 1999). This indicates that the possibility exists for a given sequence to manifest features of

different types of DNA conformation. In light of the results from the other mutant sequences, the Benzonase analysis for m23 is perhaps not as clear as the other mutant Benzonase experiments, but it is known from earlier experiments that a single CpG site is not detected by Benzonase if it lies outwith an alternating $d(CG)_n$ sequence. It should also be pointed out, that even if m23 is forming A-DNA, as it appears to be, it may be the case that it is not forming over enough DNA sequence, or at the correct region of DNA, to disrupt nucleosome positioning when incorporated into them.

In summary the mutants supply a graded response to their different sequence contexts. Overall, as the sequences are similar among the triplet mutants and as a result I do not wish to overinterprete the data. However, the combined information from structural studies, Benzonase analysis and DNaseI analysis do provide a reasonable explanation for a link between methylation induced A-form or A-like DNA and the translational positioning of nucleosomes. That is, the fully methylated CpG triplet is absolutely required to show the best reactivity with the Benzonase, DNaseI and nucleosome positioning data inclusively. The methylated mutant m3 shows the best Benzonase alignment with the wild-type digestion pattern and also shows the best displacement of nucleosome 5A. Mutant m1 shows good reactivity with Benzonase over the methylated sites but with less hypersensitivity than wild-type. It still shifts nucleosome 5A, but with poor efficiency. The methylated mutants m2, m25 and m8a also react well with Benzonase but not with the same pattern as the wild-type mutant. They also have the ability to shift nucleosome 5A but not as efficiently as the methylated LE or m3 mutants. Differences between their abilities can be partly explained by sequence context. The methylated mutant m23, although showing reactivity with Benzonase, is not capable of shifting nucleosome 5A. This may for reasons discussed above. Table 6.2 summarises the information gleaned from my experiments on the β^A-globin gene in addition to nucleosome and DNaseI data obtained by Dr. C. Davey.

The relevance of this structural information to the β^A-globin gene regulation is not presently understood. Methylation of the CpG triplet was induced artificially in these sequences *in vitro*. However, before embarking on this project Davey *et al.*, (1997) had already established that hypermethylation of the CpG triplet occurs in

embryonic erythrocytes and adult brain cells, whereas hypomethylation of the CpG triplet is observed in adult erythrocytes. Methylation at the CpG triplet is therefore developmentally regulated in a similar manner to other sites within the immediate promoter region of this gene (Davey *et al.,* 1997). It is also of importance to state that although the above information has been rationalised in terms of CpG density and sequence context, it is also relevant where the CpG triplet is positioned over the nucleosome. DNA sequence at the boundary of a nucleosome does not contribute substantially to the positioning signal (reviews: Simpson, 1991; Thoma, 1992). However, the CpG triplet in the wild-type situation that shifts nucleosome 5A, is located within the nucleosome, orientated towards the histone core and centred at 1.5 turns from the dyad axis. This is an exacting demand for the positioning of the CpG triplet and nucleosomes as other studies on the effect of methylation on translational nucleosome positioning have not shown any major effect due to methylation, but did affect histone-octamer DNA contacts (Nightingale & Wolffe, 1995).

At 1.5 turns from the dyad axis, the outward facing major groove of DNA is about 20 times wider than typical linear DNA and confers a kink in the DNA at this location (Hogan *et al.,* 1987). It is therefore expected that the corresponding minor groove would be significantly narrowed at this position. This position is also important for histone H4 binding and for H4 inducement of a kinked structure at this position (Ebralidse *et al.,* 1988; Arents & Moudrianikis, 1993). The reasons for methylation induced repositioning of nucleosome 5A were speculated on in terms of a structural transition induced by Davey *et al.,* (1997).

As stated Dr. C. Davey has since investigated the Z-form potential of the CpG triplet region but this proved not to be a viable explanation on linear DNA, only on supercoiled DNA (pers. comm.). I have here investigated the potential of CpG methylation to induce A-form DNA by using the A-DNA detecting enzyme, Benzonase. I have clearly shown that CpG methylation at the CpG triplet region induces an A-DNA conformation in the sequence, as detected by Benzonase. Further to this I have shown that over a range of mutant sequences that remove one of the three methylatable sites, the overall behaviour of the mutants is that there is a loss in their ability to move nucleosome 5A. This effect may be partly to do with the extent of A-DNA formation and partly to do with whether each mutant can position at the

nucleosome dyad axis efficiently. This may explain why mutant 23 displays some A-form potential but does not move nucleosome 5A. The methylation induced structural change in this mutant may be either not pronounced enough to deflect nucleosome 5A or avoids positioning at 1.5 turns from the dyad axis. In addition the existence of A-form DNA at this sequence also fits into what we know about DNA structure as it is wrapped around the nucleosome. As stated, the major groove is widened and the minor groove reduced at 1.5 turns from the dyad axis. This is completely the opposite structural conformation than that seen in A-form DNA. A-DNA has a typically narrowed major groove and an expanded minor groove. A nucleosome presented with an A-form structure at 1.5 turns from the dyad axis may not be able to manipulate any suitable change in this structure and may be very refractory to positioning at this site. In effect the nucleosome may be bounced along to find a more suitable positioning site as a result.

The existence of CpG triplets, the methylation thereof and the promotion of A-DNA may be a discrete and precise mechanism of altering nucleosome positioning at specific sites. Davey *et al.,* (1997) report that $(CpG)_3$ occurs at a frequency one order of magnitude greater than expected in a collected database of CpG island gene sequences. Also of note is that within this database there are only two instances of motifs longer than $(CpG)_5$, one of $(CpG)_6$ and one of $(CpG)_{12}$. Only a minor fraction of the CpG triplets are located within coding sequences (Davey *et al.,* 1997). This implies that if CpG triplets are involved in the modulation of chromatin structure they probably do so mainly in the promoters of CpG island containing genes where they may influence the binding of protein factors to DNA and/or to nucleosomes. CpG islands are usually methylated along the inactive X-chromosome, and at some imprinted genes, therefore regulation mediated by the formation of A-DNA type structures may be part of the transcriptional control mechanism at these genes. Along this stretch of the β^A-globin gene promoter, various regulatory factors bind to the promoter region. These include the transcription machinery and regulatory factors such as CTCF. The displacement of nucleosomes by this structural change may influence the binding of CTCF and Sp1, both of which bind to alternating CGs. It may also influence transcription factors, in addition to potentially binding one of the methyl-binding proteins, and recruiting

histone deacetylases and associated chromatin regulatory complexes, as described in section 1.2.

Are there any other examples of methylation induced repositioning of nucleosomes? Dr. C. Davey has since moved onto analysis of a variety of the imprinted gene regions including the *H19, Igf2 and Igf2r* human gene sequences. He has recently observed changes in nucleosome positioning at the human *H19* gene. The paternally inherited *H19* gene is silenced and the maternally inherited gene is active. A 2 kb CpG rich differentially methylated domain (DMD) located upstream of the *H19* gene is essential for gene repression. Within this, a 1.2 kb region functions specifically as a silencer element, or imprinting control region (ICR) on the methylated paternal chromosome (Brenton *et al.*, 1999; Drewell *et al.*, 2000). The activity of the DMD is modified depending on its methylation status. On the maternal chromosome it is unmethylated and probably acts as a methylation sensitive insulator, disrupting the interaction of downstream enhancers with the *Igf2* promoter (Thorvaldsen *et al.*, 1998; Hark *et al.*, 2000; Kanduri *et al.*, 2000a). The disruption of the maternal unmethylated ICR is partly due to binding the factor CTCF which has a CG rich consensus sequence. When the ICR is methylated, CTCF binding is abolished, allowing the downstream enhancers to interact with the paternal *Igf2* (Kanduri *et al.*, 2000b; Bell & Felsenfeld, 2000). Assays using HeLa cells transfected with the *H19* DMD region linked to a reporter gene shows that this region recruits MeCP2 and is therefore likely to do so on the endogenous methylated paternal allele (Drewell *et al.*, 2002).

Following on from discussions on the work presented here, I requested whether any nucleosome repositioning was observed that coincided with methylation of a CpG triplet. The H19 gene provides us with another example of this situation. A CpG triplet in this sequence has the ability to shift nucleosome positioning when it is CpG methylated (pers. comm.). In light of my observations on the CpG triplet at the β^A-globin gene promoter, Dr. C. Davey provided me with the human *H19* gene, which I performed Benzonase analysis on. Unfortunately because of delays with enzyme deliveries some of the sequence had digested from overexposure to radiolabel. Time constraints prevented me from repeating this experiment. The result is included in the Appendix, Chapter 7 (figure 7.2).

Another attempt in the pursuit of methylation induced A-form structures led me to analyse a Chromatin Immunoprecipitation (ChIP) gene sequence pulled down *in vivo* by Dr. I. Stancheva using xMeCP2 from NIH T5T cells, in the laboratory of Dr. R. Meehan. Dr. Stancheva provided me with the pull-down sequence from the ChIP assay which was found to contain an alternating $d(GC)_n$ string of nucleotides. I designed PCR primers for the sequence and prepared and purified fragments for Benzonase analysis. I then performed Benzonase analysis on these fragments a number of times but each time the results were not decisive. Although no methylation induced changes in DNA structure were observed I believe this was a failure of some part of the protocol and not a real result, as the ChIP fragment has the hallmarks of a good A-forming region within the DNA sequence. I have included this experiment as figure 7.3. in the Appendix, Chapter 7.

Having established that CpG methylation can induce the formation of A-DNA in certain sequence contexts and also correlates with one case of nucleosome repositioning, I turned my attention to some observations made by Dr. R. Meehan. In competition experiments I made some initial investigations into whether CpG methylation induced A-DNA formation can have a direct effect on the methyl binding proteins. This was based on an observation made by Dr. R. Meehan (figure 1.7). In this experiment it was found that the A-DNA forming double-stranded oligomers, poly (dI.dC) and poly d(I.C).d(I.C), which form A-DNA structures, were capable of competing successfully with CpG methylated DNA for MeCP1 complex binding. However, A-form dsRNA consisting of poly(I).poly(C) cannot compete for MeCP1 complex binding. I decided to see if similar observations could be made with the MBD domain of MeCP2. I expressed and purified the protein effectively and used it in bandshift analysis with a series of ds DNA oligomers which I specifically designed to either promote the formation of A-DNA or B-DNA. Table 5.1 provides the sequence of each oligomer used.

The results of these experiments show that the MeCP2 MBD domain is capable of binding both the A-dnaM oligomer (figure 5.3) containing two non-adjacent meCpGs and the B-gloM oligomer (figure 5.9) containing three adjacent alternating meCpGs. The MBD protein binds equally well to both of these oligomers. It was already known that the MBD could recognise a single methyl CpG (Nan *et al.*,

1993). From bandshift analysis here it is demonstrated that the MBD can bind to meCpGs in isolation, irregardless of the sequence context and also to a stretch of alternating $(CG)_3$ nucleotides containing three methylatable CpGs. It is therefore likely that the MBD domain operates independently of whether the target methylation site is in an A-form structure or not. This is also verified by the publication of the solution structure for the MBD domain (Wakefield *et al.*, 1999). However there are also contacts within the MeCP2 protein surrounding the MBD that may influence further what sites the protein can bind to, and of course, other chromatin associated proteins may additionally influence what subset of meCpG sites the MeCP2 protein can bind to.

Bandshift experiments using an unmethylated sequence that forms A-DNA also shows interaction with the MBD protein (figure 5.4, 5.7). The bands formed are specific, but only at high concentrations of the MBD. When competitor DNA was added to the situation the A-dnaU oligomer interaction is diminished (figure 5.5). When the MBD protein is given a B-form DNA substrate as a potential binding sequence, the result is a smear across the gel. There are no specific contacts in this situation (figure 5.6). Another unmethylated oligomer, B-gloU also shows a smear across the gel with the MBD protein (figure 5.8). This sequence corresponds to the CpG triplet region of the β^A-globin gene analysed in Chapter 4 and may potentially form Z-DNA in the unmethylated state (see table 1.2). Following on from these initial experiments on the MBD protein, I then used purified xMBD1 protein to determine which oligomers the protein could recognise. The xMBD1 was provided with the A-dnaU and A-dnaM oligomers as potential substrates. The protein seems to bind equally well to both oligomers. There does not seem to be any enhanced reactivity with the methylated oligomer, A-dnaM, unlike the way the MBD protein reacts with this sequence. This may be for several reasons. Perhaps the xMBD1 protein can recognise A-form DNA and as a result does not interact to any greater extent with the methylated oligomer. Of course it is also possible that the xMBD1 protein interacts with A-form DNA minimally but cannot interact with the two methylation sites presented in the A-dnaM oligo well. Although the MBD from human MBD1 can bind to a single methylated site, this may differ with the *Xenopus* form, despite the overall sequence similarity.

These initial bandshift experiments perhaps provide some information with respect to the role of MeCP binding to the overall mechanism of methyl-mediated repression. The MBD of MeCP2 and the xMBD1 proteins recognise at least one methyl CpG. The MBD seems capable of recognising the meCpG in sequences that potentiate A-form DNA and also in sequences that do not. I would conclude that although both proteins show interactions with A-form DNA, this is not as specific as the effect seen with the MeCP1 complex, where the presence of A-forming sequences efficiently competes with methylated DNA for MeCP1 complex binding (figure 1.7). The MeCP1 complex has been shown to require a densely methylated substrate of 11 or more methyl CpGs (Meehan *et al.,* 1989). A long alternating $d(CG)_n$ substrate could provide a very good sequence for A-DNA formation and this may be important in the decision of what sites the MeCP1 complex can bind to, leaving other methyl binding proteins such as MBD1 and MeCP2 to recognise and bind to other methylated sites. However, the meCpG binding activity in the MeCP1 complex has been shown to be due to MBD2, which binds to a single meCpG (Ng *et al.,* 1999). Why then does this discrepancy exist? It is very possible that other proteins within the MeCP1 complex direct MBD2 to meCpG regions that are densely methylated. It is therefore also possible that such proteins associated with the MBD1, MeCP2 and other MeCPs, perhaps do so on the basis of methylation density and resultant secondary structure alterations.

6.3 How can methylation induced changes in DNA structure influence where and when methyl-binding proteins bind?

In this thesis there is evidence that CpG methylation can induce potential A-form regions in DNA as detected by Benzonase analysis. It is also known that different MeCPs can detect and bind to different meCpG densities and are influenced by sequence context. How CpG methylation density may relate to changes in DNA structure and subsequent protein binding is illustrated in figure 6.2 below.

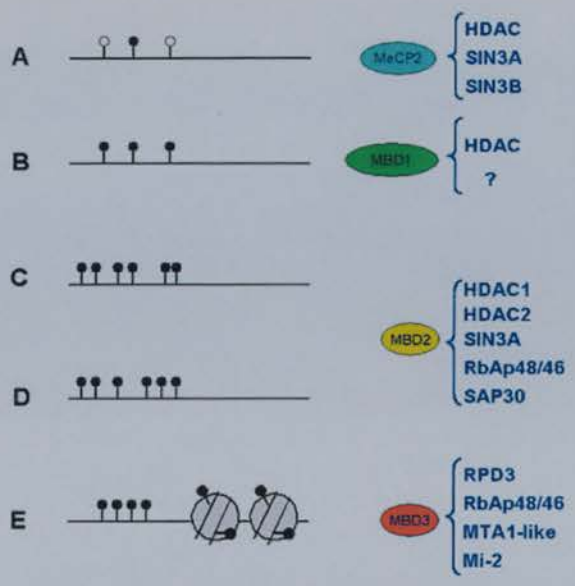


Figure 6.2 MeCPs target corepressor complexes to methylated DNA. On the left are different situations the MeCPs may recognise. Empty circles represent unmethylated CpGs, whereas full circles are methylated CpGs. (A), hypomethylated DNA with occasional meCpGs. (B), fully methylated sequence with a low density of CpGs. (C) and (D) are two sequences with a high number of CpGs but with different organisations that may be recognised by different MeCPs. (E) includes the nucleosome structure to illustrate that it may be the ability or positioning of methylated CpGs around the nucleosome that can influence the binding of MeCPs and potentially other components of the chromatin network. (Image from Ballestar & Wolffe,2001).

Ballestar and Wolffe (2001), state that different sequences containing different CpG densities have different affinities for MeCP proteins. Their work pertaining this remains unpublished. They also speculate on the role of DNA structure in this process. I believe in this thesis a potential mechanism is revealed. If methylation induced A-DNA formation is not for the recognition of MeCP proteins, or some component of an MeCP-associated complex, it must have some other potential role. The first of these could be in the direct exclusion of nucleosomes. It is recognised that some DNA binding factors require removal or remodelling of nucleosomes, to gain access to their cognate binding sites (diMauro *et al.,* 2002). MeCP2 for example, does not require this. Overexpression of MeCP2 in transient transfection studies shows that the MeCP2 molecules can still gain access to and bind mouse heterochromatin (Nan *et al.,* 1996). It must be able to gain direct access to binding sites. One explanation is that MeCP2 can bind nucleosomal DNA to form discrete complexes (Chandler, 1999). Another explanation offered here is that nucleosomes encountering A-DNA stretches, particularly of the type $d(CG)_3$ are moved away from this region leaving it free to interact with MeCP2. These two explanations are not

mutually exclusive. Both could exist for different situations of MeCP binding. Further to this explanation, it has been shown that bending and kinking disruptions to DNA, and particularly the formation of A-DNA, can disrupt or prevent nucleosome formation (Hovatter & Martinson, 1987). Further to this, it has also recently been speculated on that the formation of both positive and negative superhelical writhes are organisers of local chromatin structure for transcription in eukaryotes (review: Ohyama, 2001). Therefore in regions with a high density of methylated CpGs, nucleosomes may be refracted and MeCPs allowed access to their binding sites. We have also observed that in the case of the β^A-globin promoter, CpG methylation of a CpG triplet sequence can alter nucleosome positioning. This too could affect the binding of methyl-binding proteins and/or alter the positioning of chromatin-associated proteins on DNA.

The solution structures of the unliganded MBD domain from human MBD1 (Ohki *et al.,* 2001) and rat MeCP2 (Wakefield *et al.,* 1999) provide additional pieces of information to the puzzle of methylation mediated repression mechanisms. Both MBD domains have similar structures suggesting that all MBDs adopt a similar structural fold (figure 1.9). The MBD domain is wedge shaped, the faces of which are comprised of a β-sheet on one side and an α-helix and hairpin loop on the other. A highly flexible loop connects two of the β-strands at the vertex of the wedge. When the MBD domain of MBD1 binds to DNA, the flexible loop undergoes a major structural change, adopting a hairpin structure consistent with an induced fit mechanism. Unusually, the recognition surface of MBD1 is asymmetric. Therefore, each methyl group can be recognised independently. One methyl group fits into one pocket of the hydrophobic contact surface, the other methyl group fitting into another region of the hydrophobic surface. Residues from Loop 1 of the structure contact one side of the DNA backbone, while residues from the α-helix contact the opposite strand backbone. Because of the asymmetry of binding to a symmetrically methylated CpG, the orientation of the MBD and any associated protein complex, will change by 180° depending on which functional groups bind to a meCpG on a particular strand (Ohki *et al.,* 2001, Wade & Wolffe, 2001). The functional consequences of asymmetric binding remain unexplored at present. However, this may also mean that as the distortion of a particular meCpG increases, due to DNA

bending, the formation of alternative DNA conformations such as A-DNA and Z-DNA, the greater the influence on the asymmetric binding of MeCP associated protein complexes. Again, it is unknown what influence an altered DNA structure has on the 3-dimensional position of methyl associated complexes, and what this might mean in a functional context.

An intriguing possibility for the role of A-form DNA in the regulation of methylation-mediated repression, is the potential involvement of RNA in the process. RNA mediated gene silencing involves small RNAs that can trigger gene silencing in both the cytoplasm and nucleus. RNA mediated silencing can occur at the transcriptional and post-transcriptional level. In brief, the recently discovered process of RNA interferance (RNAi) in animals, (Fire *et al.,* 1998), and the related phenomenon of post-transcriptional gene silencing (PTGS) in plants involve a dsRNA that is cleaved into shorter units that act as guides for recognition and cleavage of homologous mRNA. The dsRNA guides can be created by transcription through inverted DNA repeats, simultaneous synthesis of sense and antisense RNAs, viral replication and the activity of cellular or viral RNA dependent RNA polymerases (RdRP) on single-stranded templates (review: Matzke *et al.,* 2001). Similarly, RNA-directed DNA methylation in plants requires dsRNAs that are cleaved into small RNAs similar to the guide RNAs formed in the PTGS/RNAi processes. Only DNA sequences complementary to the guide RNA become modified, suggesting RNA-DNA interactions, a highly specific interaction, but with the possibility of any sequence being a potential target of this mechanism.

How might this mechanism be used to target methylation in mammals? The protein machinery involved is currently not known. There are suggestions that small RNAs form DNA-RNA hybrids in unusual structures that might attract Methyltransferases (Mtases), (Mette *et al.,* 2000; Waterhouse *et al.,* 2001; Smith *et al.,* 1991). Alternatively, Mette *et al.,* suggest that small RNAs may bind Mtases and thus guide them to target DNA sites. For example an MTase from plants possesses a chromodomain and is called a chromomethylase (Lindroth *et al.,* 2001). Chromodomains are typically involved in interacting with chromatin proteins to modulate gene expression. A chromomethylase has not been discovered in mammals thus far. In addition a selection of dsRNA deaminase cDNAs from *Xenopus,* thought

to be involved in RNA editing, possess a C-terminal region with homology to C5-Methyltransferases (Hough & Bass, 1997). Despite the fact that this provides an astonishing ancient link between methylation-mediated repression mechanisms and our ancient origins in an RNA world, it also raises the possibility that MTases may still be able to recognise the A-conformation in DNA. However, adding weight to an RNA-chromatin driven hypothesis, is the *Drosophila* histone acetylase MOF, which contains a chromodomain shown to interact with RNA (Akhtar *et al.,* 2000). It is therefore possible that not only DNA methylation, but chromatin modifications such as acetylation and methylation of histones might also be targeted by RNA. In mammals, X-chromosome inactivation and some cases of genomic imprinting in mammals involve non-coding RNAs or overlapping sense and antisense RNAs, which may play a role in a homology driven mechanism for DNA methylation and/or chromatin modifications.

A role for A-form DNA in RNA mediated repression? Thus far I have proven the existence of CpG methylation induced alternative structures in DNA. The structures formed are akin to the A-conformation of DNA, as detected by Benzonase. The extent of transition to this form is dependent on the number of methylatable sites and the sequence context of the DNA being methylated. It is possible that these sequences are differentially recognised by methyl binding proteins that have recognition regions apart from their MBD that are sensitive to the helical conformation of DNA. However, preliminary bandshift analyses with purified proteins only show slight preferences for A-form DNA, and I therefore do not wish to conclude that they specifically recognise A-DNA, although many remain untested. If the MeCP protein alone does not recognise the change in structure, it may require another component.

There is preliminary evidence for this from unpublished competitive bandshift analysis kindly provided by Dr. R. Meehan, that this might be the case for the MeCP1 complex. The complex is successfully competed away from methylated DNA with an excess of A-forming non-methylated dsDNA sequence. The complex may therefore contain an uncharacterised component that is competed away from its methylated target sequence by an excess of A-forming sequences. It is also possible that an MBD within the complex or another associated protein, may independently

recognise A-form structures and as a result, recognise and bind to the excess of A-forming structures in the competitive bandshift. This explanation would also account for the fact that the MeCP1 complex requires a long alternating methylated binding sequence, even though the MBD2 protein which forms an integral part of the complex, can bind singly methylated CpGs.

We also know that the formation of A-DNA in methylated sequences is refractory to nucleosome positioning from unpublished collaborative data supplied by Dr. C. Davey. This is especially true if the methylated sequence coincides with regions of nucleosome induced bending in the DNA. This may have consequences in terms of higher order chromatin structure directly by the displacement of nucleosomes, or may simply allow methyl binding proteins access to their binding sites.

Finally the aforementioned possibility of an RNA component is intriguing, as dsRNA is known to possess an A-conformation. Unlike DNA, a dsRNA always adopts an A-form helix (Dock-Bregeon *et al.*, 1989). Despite the fact that A-form dsRNA does not compete for MeCP1 complex binding (figure 1.7) I am confident that some role for dsRNA remains to be discovered in humans. Perhaps A-form DNA/RNA hybrids could compete for complex binding? This conformational change may promote homology driven binding by complementary A-forming RNAs that recognise the helical change as well as the sequence. There is some evidence for this from *Xenopus laevis.* An unusual dsRNA binding protein has been discovered called dsRBP-Zfa. It is unusual in that it contains zinc fingers as the recognition domain, a highly common motif for transcription factors and other DNA binding motifs, and also possesses a nuclear localisation signal (NLS), (Finerty & Bass, 1999). This protein binds to dsRNA and also to RNA-DNA hybrids with almost equal affinity, via separate zinc finger motifs. Other dsRNA binding proteins have reduced ability to bind RNA-DNA hybrids. There are separate zinc fingers for DNA and RNA interactions. Crucially this protein greatly prefers to recognise the A-conformation, independent of sequence context and is the first Zinc Finger protein that shows specific binding to dsRNA and RNA/DNA hybrids (Finerty & Bass, 1997).

The possibility exists that similar dsRNA binding proteins with or without sequence preference can interact with dsRNA/DNA sequences based on an affinity

for A-form structures. In speculation, dsRNA binding proteins may associate with small RNAs produced from non-coding sequences associated with regions of DNA targeted for methylation. They may associate with complexes that contain MTases that have RNA binding abilities, or indeed with histone acetylases that have RNA binding abilities. These RNA associated protein complexes may contain helicases to unwind DNA and components of the methylation machinery, such that dsRNA binding proteins, Methyl-binding proteins, and other components are then tethered to the correct region of DNA destined for targeted DNA methylation and/or associated chromatin modifications.

No evidence exists thus far in mammals for a mechanism based on the experimental evidence presented in this thesis. A role for the formation of A-form DNA in methylated gene sequences has also not been elucidated thus far. I believe that I have presented evidence to support a potential role for A-form DNA in the mechanism of methyl-mediated repression. The elucidation of this mechanism undoubtedly awaits more surprises. It is a process that is becoming intriguingly complex, with plenty of twists and turns yet to be discovered.

Chapter 7: Appendix

This chapter describes work that was carried out in the process of investigating the CpG methylated DNA structure. Some of this work concerns the Benzonase assay and consists of data not shown in the main text. The rest refers to the use of a retroviral integrase based assay to determine regions of non B-form DNA structure. This assay has been described in Chapter 3 and illustrated in figure 3.1.

7.1 Benzonase analysis of a DNA fragment from pBR322

A DNA fragment corresponding to the reverse strand of the pBR322 fragment analysed in Chapter 3 (see figures 3.4 & 3.5) was analysed by Benzonase analysis. The resultant gel from this experiment is detailed here as the results were perhaps not clear enough to incorporate in the main thesis. The aim of this experiment was to show that the same fragment could display methylation induced structural changes on the reverse strand at the same position observed on the forward strand. This effect was later proved by analysis of the β-globin gene fragment shown in figure 4.4, which confirmed that methylation induced structural changes were observable on the reverse strand and thus more likely to be a change in structure on both strands and not restricted to one of the DNA strands. The results of Benzonase analysis of the reverse strand pBR322 fragment are shown in figure 7.1

7.2 Benzonase analysis of a DNA fragment from the human H19 gene.

An attempt was made to show that a methylation induced structural change at a CpG triplet region in the human imprinted H19 gene, could react with Benzonase to show enhanced cleavage at this region. The methylated CpG triplet region has been shown to displace a nucleosome by Dr. C. Davey in the H19 gene. If Benzonase analysis can also highlight this region as a potential A-forming sequence, then it would provide the second example of a methylated CpG triplet that can directly alter nucleosome positioning at a region of potential importance in terms of gene regulation. Unfortunately, degradation of this sequence ensued due to time delays in receiving restriction enzymes. The results of this experiment are shown in figure 7.2.

The H19 gene, as well as other imprinting genes, are worth investigating in the future in terms of potential regions of methylation induced A-form DNA structures coupled to nucleosome rearrangement and this information is provided here as a good candidate for investigation by others. Time constraints prevented me from investigating further.

7.3 Benzonase analysis of a DNA fragment from a xMeCP2 ChIP assay.

Another attempt to show methylation induced A-form structures was performed on a fragment pulled down in a Chromatin Immuno-Precipitation assay with xMeCP2 as bait in NIHT5T cells. The DNA sequence pulled down by xMeCP2 is called Chip30 and was cloned into pBS-SK(-). The plasmid insert was amplified by PCR and subjected to Benzonase analysis. The resultant gel is shown here in figure 7.3, but the sequence is so long that banding patterns are not clear on this gel and the results inconclusive. However, the sequence is probably worth reinvestigating by other interested parties in the future and is included here for reference purposes.

7.4 The integrase assay: Expression of HIV-1 integrase

Initial attempts to investigate the structure of methylated DNA were based on the use of an assay based on the ability of HIV-1 integrase to direct integration of dsDNA sequences containing viral recognition sequences at each end, into a target DNA sequence. The assay is described in section 3.1.1 and figure 3.1. I attempted to use this assay in the investigation of CpG methylated DNA structure. The process began with the purification of HIV-1 integrase by essentially the same methods as that described for the histidine-tagged MBD protein. The HIV-1 integrase protein was expressed from a pET6h plasmid vector in *E. coli* BL21(DE3)pLysS cells. The protein extracts containing HIV-1 integrase were purified on a cation exchange column and further purified on a Nickel charged affinity column. Purified protein was subsequently treated with Thrombin protease to remove the histidine tag. Thrombin was removed by purification over a Benzamidine 6B column that has a strong affinity for thrombin, and subsequently re-purified over a Nickel charged column to remove excess cleaved histidine tags from the protein solution. The purified protein was stored in a glycerol-based suspension and stored at -80^0C.

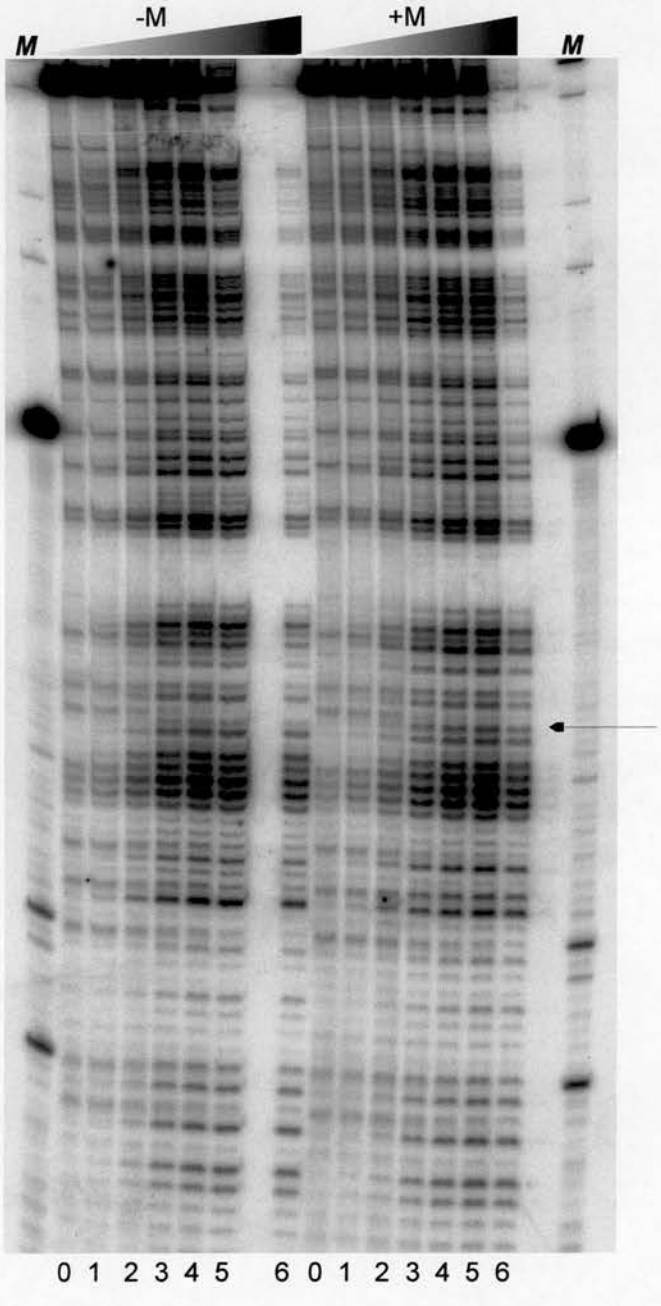


Figure 7.1 Benzonase Analysis of the reverse strand of a pBR322 fragment. PAGE/Urea analysis of the The forward strand of this fragment has already been analysed in figure 3.5. The reverse strand was also Benzonase digested to demonstrate that the methylation enhanced cleavage observed on the Benzonase digested forward strand was also present on the reverse strand. There is a site in the sequence which may or may not be due to methylation enhanced cleavage. This is marked with a long arrow. There may also be some DNaseI or radiolabel caused digestion in this fragment prior to Benzonase analysis, as the banding pattern shifts on addition of Benzonase. This gel is incorporated for reference purposes. The existence of Benzonase enhanced cleavage on the reverse strand of methylated fragments was later proved by experiments in Chapter 4 (see figure 4.4). **-M,** unmethylated DNA. **+M,** methylated DNA. ***M,*** marker lane containing forward strand of pBR322 fragment digested with *Hha*I.

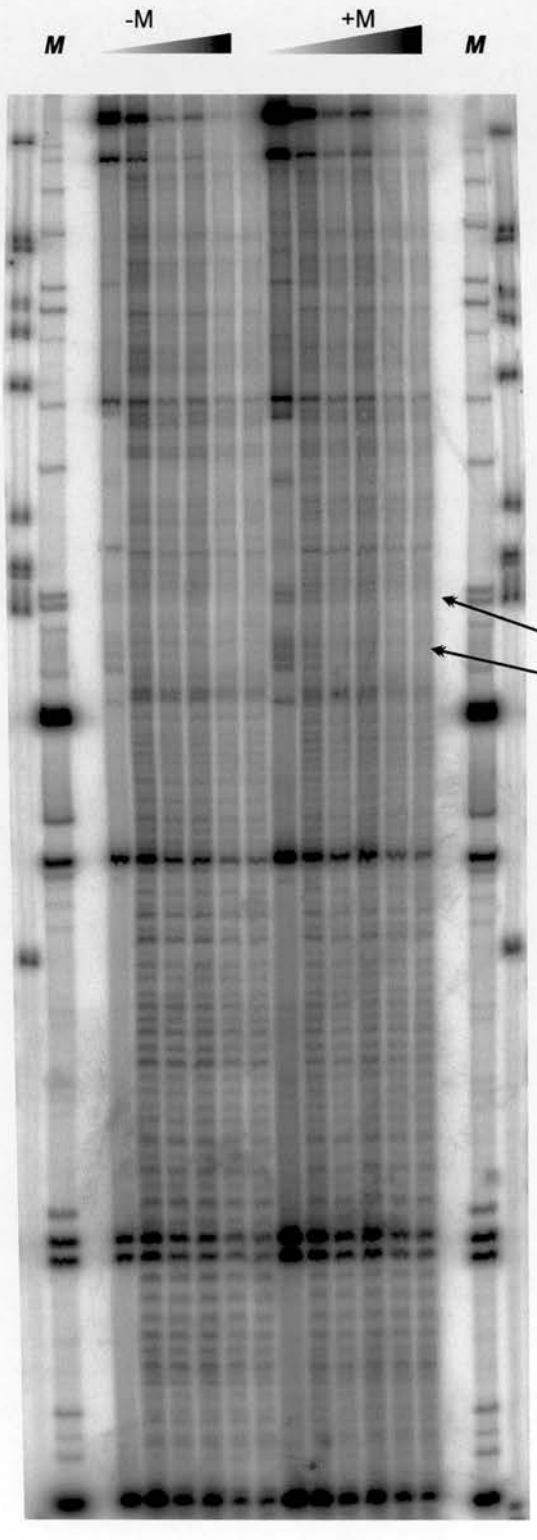


Figure 7.2 Benzonase analysis of a region of the human H19 gene. The human H19 gene was analysed by Benzonase analysis and the results run on a 8 PAGE/Urea gel. It contains a CpG triplet region within the Differentially methylated Region (DMR), a region of importance in terms of gene expression. There has been some digestion of this fragment due to long waits for restriction enzyme delivery. -M, unmethylated. +M, methylated. *M,* marker lane. H19 gene fragment digested with *Hha*I. This gel is provided for referenve purposes and also for information pertaining to the H19 gene. I believe re-analysis of this sequence by others may reveal a similar phenomenon to that observed at the β-globin promoter region in terms of nucleosome displacement.

Arrows indicate regions of potential methylation induced structural change as detected by Benzonase.

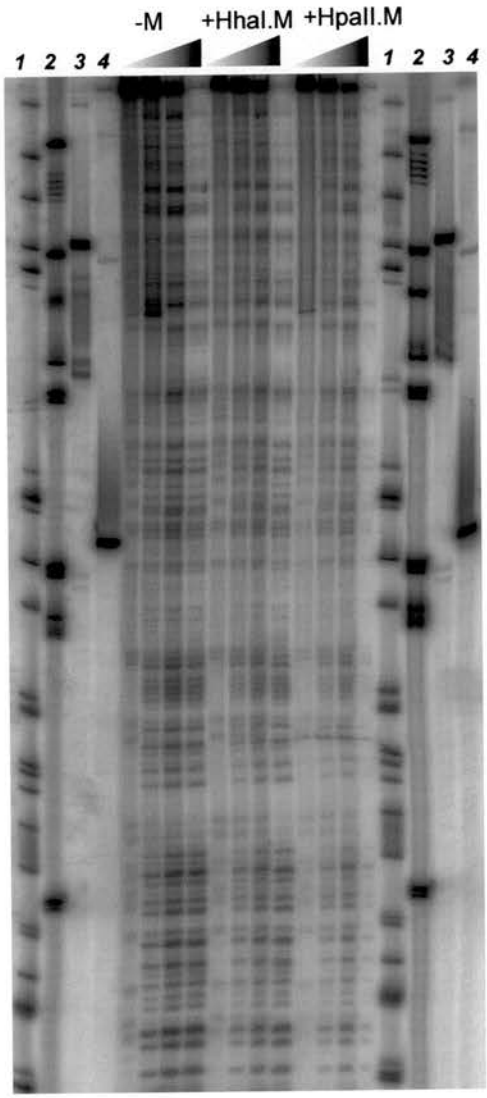


Figure 7.3 Benzonase analysis of a Chromatin Immunoprecipitation assay DNA sequence. The DNA sequence used here was pulled down by xMeCP2 in a ChIP assay to determine the preferred binding site of xMeCP2 protein *in vivo*. The sequence was pulled out on several occasions. It is referred to as ChIp30 as this is the simply the clone number. The sequence contains a CpG triplet sequence like that analysed in the b-globin gene promoter region (Chapter 4). ***1,*** pBR322/BsuR1 marker. ***2,*** pUC19/MspI marker. ***3,*** *Hha*I digested DNA fragment. ***4,*** *Hpa*II digested DNA fragment. Samples were either unmethylated, -M or methylated with either *Hha*I methylase (+HhaI.M) or with HpaII methylase (+HpaII.M). Samples were Benzonase digested and run on a 6 % PAGE/Urea gel and autoradiographed.

Other aspects of the assay involved the preparation of target DNA for integration of DNA mediated by the purified enzyme. The pUC9-α_2-globin sequence was used for this purpose. The plasmid was transformed into XL1-blue cells and midipreped. The plasmid was cleaned twice by phenol-chloroform extraction and ethanol precipitation. DNA primers were designed for the purposes of the assay. The forward primer was designed to anneal to the DNA substrate destined for plasmid integration. The reverse primer was designed to anneal to a region just outside the gene insert of the pUC9 plasmid. The DNA substrate was a short dsDNA fragment containing recognition sites for HIV-1 integrase at each end. Initial attempts to assemble all the components of the assay did not appear to work. Target plasmid DNA along with purified HIV-1 integrase and substrate DNA were incubated together, according to the method of Chow, 1997. After the integration reaction, PCR was performed to identify regions of DNA within the α_2-globin sequence containing integrants. The forward PCR primer was radiolabelled with ^{32}P-γ-ATP. The reverse primer remained unlabelled. Primers were added to the plasmid mix, which should contain HIV-1 directed DNA integrants. PCR was performed on the target plasmid. Fractions were taken and run on a 5% PAGE gel to identify if there was a ladder of radiolabelled DNA fragments produced, prior to separation by denaturing PAGE/Urea. There did not appear to be a ladder or DNA smear produced as determined by PAGE. A denaturing PAGE/Urea gel was then run to determine more clearly if there were any PCR fragments produced. The experiment did not seem to work. Many attempts were made but to no avail. To ascertain that the plasmid did indeed contain the correct sequence it was subcloned into a pBS-SK(-) plasmid and sent away for sequencing within the University. The resultant gene sequence confirmed that the plasmid did contain the full sequence of the HIV-integrase protein. Primer sequences were re-evaluated and the procedure re-tried. Experiments to discover the stage of the procedure were also carried out. For example, an experiment was carried out to discover whether the substrate DNA was being efficiently integrated into the target sequence or not. Variations in the PCR conditions used were attempted. Although the experiments to decipher the reasons for failure of the assay were not exhaustive, they were very time consuming and cumbersome. It was concluded that perhaps there was a problem with the extensive

rounds of purification the HIV-1 integrase protein had suffered, that perhaps it was not folded correctly, or for some unknown reason it was not active in my attempts at the assay. The results of the assay, had it worked correctly, were to discover regions of methylation induced structural change as detected by HIV-1 integration hotspots. The nature of the structural changes could not be decided on from this analysis. It was hoped also to use a chiral probe to detect the formation of A-form DNA in these sequences, but requests for the Ruthenium based complex were refused. Because of the cumbersome nature of this approach, the problems associated with optimisation, the fact that it would not be easily determinable if any methylation-induced structures were of the A-conformation, this approach was abandoned and the more decisive nature of Benzonase explored for structural analysis of methylated DNA.

7.5 Expression and Purification of the HIV-1 integrase protein.

A plasmid obtained from Dr. Craigie containing a soluble active mutant of the HIV-1 integrase protein (Jenkins *et al.,* 1996) was expressed and purified as described in Chapter 2. In brief, cells were transformed with pIN^{1-288}/F185K/C280S, were grown and protein expression induced with IPTG. Cells were harvested and homogenised to release cellular proteins. Cell extracts were centrifuged to remove cell debris and filtered to ensure as much cellular material as possible is removed. Purified extracts were applied to a Nickel charged Chelating Sepharose Fast Flow (Pharmacia) column as there is a strong binding reaction with Histidine tags and Ni^{2+} ions. The protein was initially eluted with Imidazole, a strandard elutant for Ni^{2+} charged elution (data not shown). Although this elution worked very well, there was some degradation of the protein. It was also feared that the high concentrations of Imidazole used were interfering in some way with further purifications, the purity of the protein and the ability of it to perform in the Integrase assay. L-histidine was then used instead of Imidazole. It is a successful competitor at low concentrations and results in a highly purified product with high reactivity. Data for the expression and L-histidine purification of HIV-1 integrase is shown in figures 7.4 & 7.5. The protein concentration was estimated by spectrophotometry to be approximately 1.2 mg/5.5 ml or 0.21 mg/ml and migrates at approximately 33 kDa.

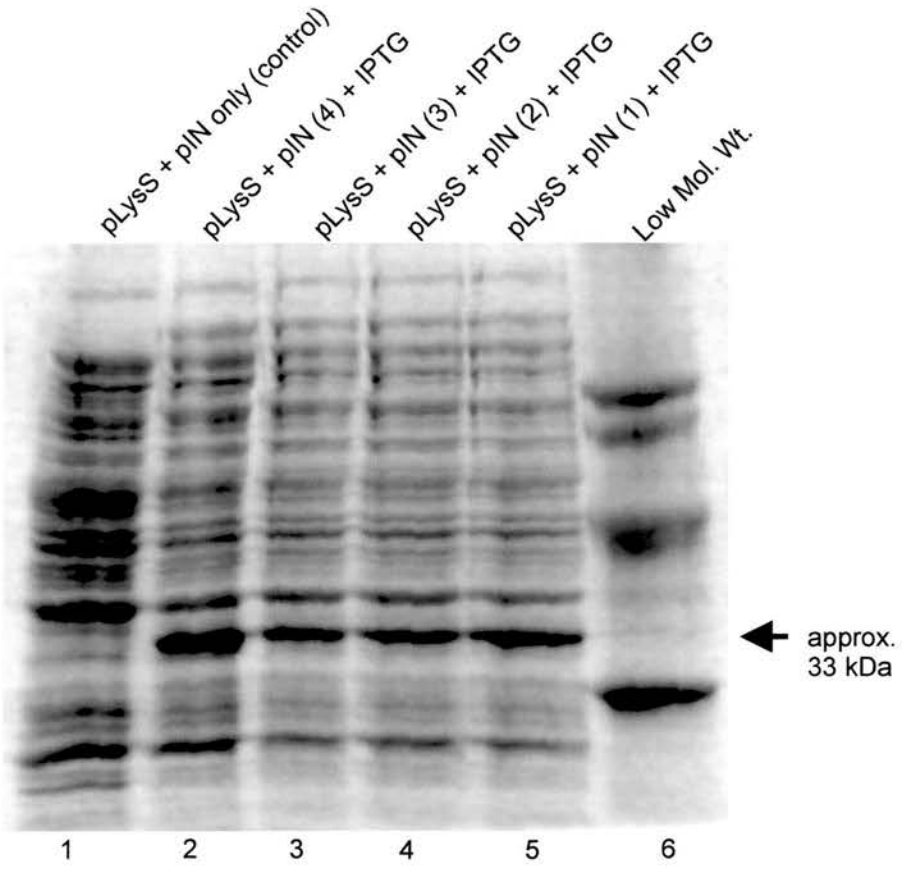


Figure 7.4 Expression of HIV-1 integrase protein. Image is an 8 % SDS-PAGE gel. A plasmid containing a soluble active mutant of the HIV-integrase protein was transformed into *E. coli* BL21(DE3)pLysS cells. Batches of cell cultures were induced with IPTG apart from lane 1, which contains a cell culture not induced with IPTG and therefore contains only *E. coli* proteins. The arrow indicates the position of HIV-1 integrase expression in the cell culture.

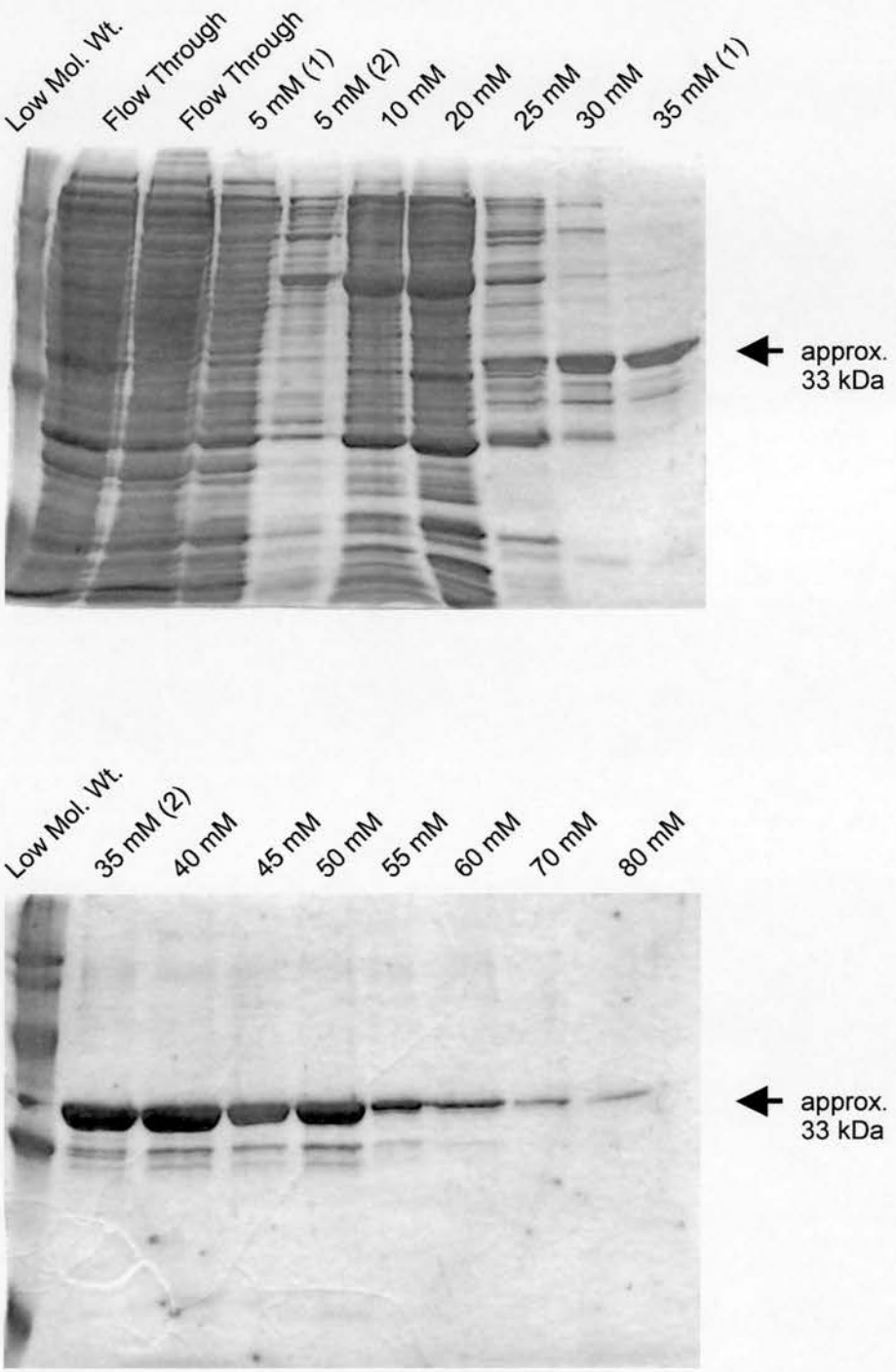


Figure 7.5 Purification of a histidine tagged HIV-integrase protein. The soluble HIV-1 integrase mutant expressed and shown in figure 7.4 was applied to a Nickel affinity column and eluted with increasing concentrations of L-histidine. A closed arrowhead indicates the position of the eluted protein. There is a small amount of degradation in the Integrase purification, but this is minimal in comparison to the concentration of purified intact protein.Image is an 8 % SDS-PAGE gel.

7.6 Cleavage of HIV-1 integrase histidine tag with Thrombin

Some proteins require cleavage of their artificial tag to enable correct folding of the protein for proper enzymatic activity. It was decided that the presence of the histidine tag might inhibit the activity of HIV-1 integrase in the Integrase based assay. Thrombin protease was used to cleave at two thrombin recognition sites to remove the histidine tag as follows. Combined samples from elution with 40 mM and 45 mM L-histidine (figure 7.5) were dialysed to reduce the L-histidine concentration and re-applied to a Nickel charged Chelating Sepharose Fast Flow (Pharmacia) column. Thrombin protease (Pharmacia) was diluted as required and 400 units applied to the Nickel-bound HIV-1 integrase protein. The column was stirred gently and incubated at room temperature for two hours and the cleaved protein washed through the column with elution buffer F containing only 20 mM L-histidine. Histidine tags should remain bound to the column and the flowthrough mostly contain cleaved protein. The flowthrough appeared to contain cleaved protein (data not shown) and was applied to a Benzamidine Sepharose 6B column to remove the contaminating thrombin protease.

7.7 Removal of Thrombin with a Benzamidine Sepharose 6B column

Eluted cleaved protein was applied to a Benzamidine Sepharose 6B (Pharmacia) column to remove thrombin protease. Benzamidine Sepharose is p-aminobenzamidine (PAB) covalently attached to Sepharose 6B. PAB is a synthetic inhibitor of trypsin-like serine proteases and can thus be used for their removal. Cleaved protein was applied to a Benzamidine column in a solution containing 200 mM NaCl. The column was washed through with binding buffer (50 mM Tris-Cl, pH 8.0 containing 200 mM NaCl), to remove the cleaved protein. Thrombin was later eluted in separate fractions with 500-1 M NaCl. The results of this experiment are shown in figure 7.6.

7.8 Concentration of eluted protein on a Fractogel EMD-SO_3^- column.

Fractions containing cleaved and purified HIV-1 integrase were combined and concentrated by application to a Fractogel EMD-SO_3^- strong cation exchange column. Proteins are reatained efficiently on the column when the pH of the buffer is

about 1 unit below their isoelectric point (pI). Samples were loaded in a 200 mM NaCl containing buffer and washed to remove any contaminants. Protein was eluted sharply in small volumes using 1 M NaCl. The results are shown in figure 7.7.

7.9 Western Blotting to detect presence of HIV-1 integrase

Western Blotting was performed at different stages of the purification procedure. Two examples are included here. Figure 7.8 clearly shows the reaction to an anti-histidine tag antibody. Only cells that contain the expressing plasmid that have also been induced with IPTG show strong reactivity with the antibody, confirming that the protein is indeed a histidine tagged product produced by IPTG induced expression from the transformed plasmid vector.

7.10 The Integrase Assay: Attempts at DNA structural analysis.

Having gone to great lengths to purify the HIV-1 integrase protein, I assembled the other components for the Integrase assay. A pUC9 plasmid containing the human α_2-globin gene was midipreped and purified. Approximately 10 µg of plasmid DNA was methylated with *Sss*I methylase. Unmethylated and methylated plasmids would provide target DNA in the assay. Donor DNA, destined for integration to the plasmid DNA was prepared in accordance with Chow, 1997. The oligomers used were H-U5V2 (5'-ACTGCTAGAGATTTTCCACAT-3') and H-U5V1-2 (5'-ATGTGGAAAATCTCTAGCA-3'). 100 pmol of H-U5V1-2 was 5' labelled with ^{32}P-γ-ATP and annealed to H-U5V2. This dsDNA labelled fragment served as the Donor DNA. Attempts at the Integrase assay ensued. 5 pmoles of HIV-1 integrase were incubated with 1µg of Target DNA plasmid and 15 nM of Donor dsDNA oligomers. PCR was carried out on the fragments but it did not produce a smear or ladder of DNA products as hoped. Samples of several PCR reactions were run on 12 % PAGE gels, but not products were observed (data not shown). At this point it was attempted to see if the HIV-1 integrase was capable of integrating the dsDNA oligomers. 12 % PAGE gels were again run, but directly after the integration reaction, prior to the PCR step. Again, the plasmid did not appear to show strong radioactivity as assessed by autoradiography, and there was assumed to be little or no incorporation of Donor DNA into target sequence (data not shown).

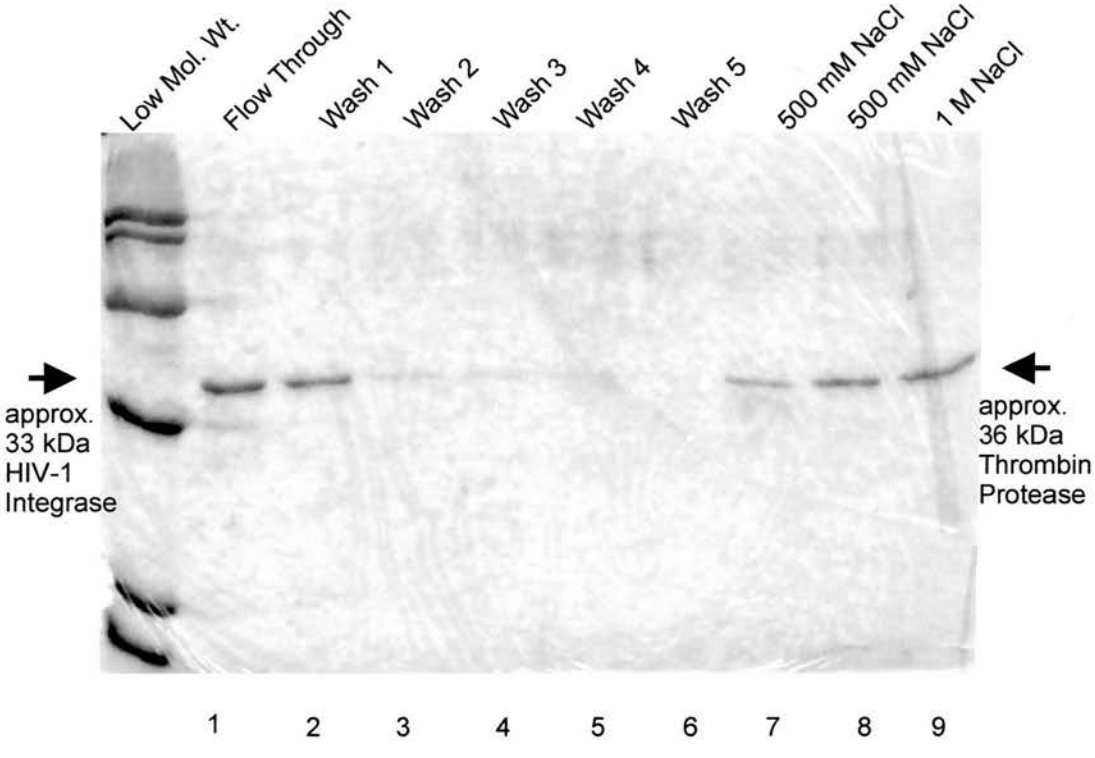


Figure 7.6 Removal of Thrombin Protease with a Benzamidine 6B column. The cleaved HIV-1 integrase protein is applied to the column in solution with Thrombin Protease. Thrombin binds to the column with strong affinity and HIV-1 integrase is washed through in the flow through, lane 1, and the first few washes, lanes 2 - 4. Washing the column with a high salt solution removes Thrombin from the column; lanes 7 - 9. Both HIV-1 integrase and Thrombin protease have similar molecular weights and are therefore not readily distinguishable by gel position, but are distinguishable by elution properties. Image is an 8 % SDS-PAGE gel.

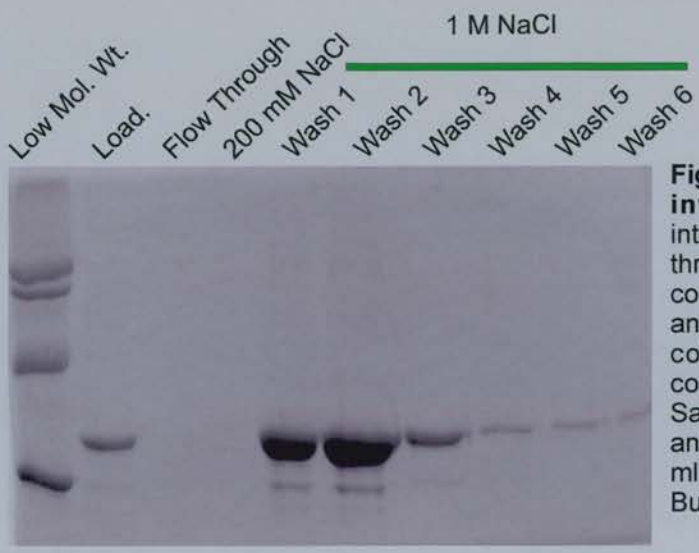


Figure 7.7 Concentration of HIV-1 integrase. Fractions of HIV-1 integrase that have been purified, thrombin cleaved and separated from contaminating thrombin were pooled and applied to a cation exchange column, EMD-SO3- (Merck) to concentrate the protein samples. Samples were loaded in 200 mM salt and eluted in concentrated form with 2 ml volumes of 1M NaCl containing Buffer solution.

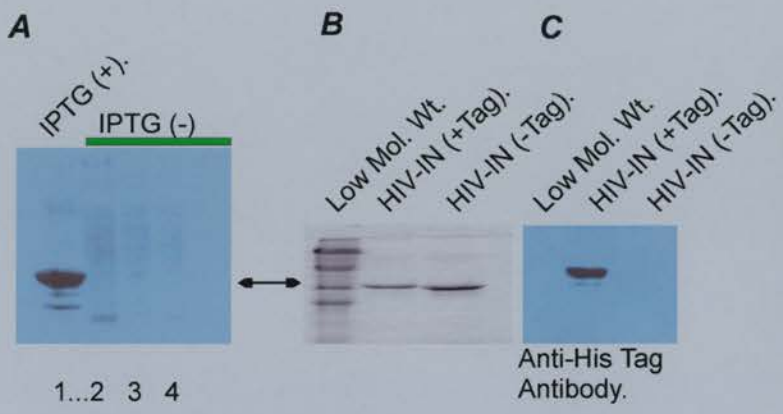


Figure 7.8 HIV-1 integrase expression. ***A,*** Western blot of an IPTG induced cell culture (+) and three other cultures without IPTG induction (-). **1,** cells only. **2, 3, 4,** cells with pIN plasmid. ***B,*** Gel used in Western blot, pre-transfer. HIV-integrase expression (HIV-IN) with (+) or without (-) a histidine tag. ***C,*** Western blot of gel image shown in ***B,*** confirming that the histidine tag has been removed. The antibody used in ***A & C,*** is a monoclonal anti-histidine tagged antibody.

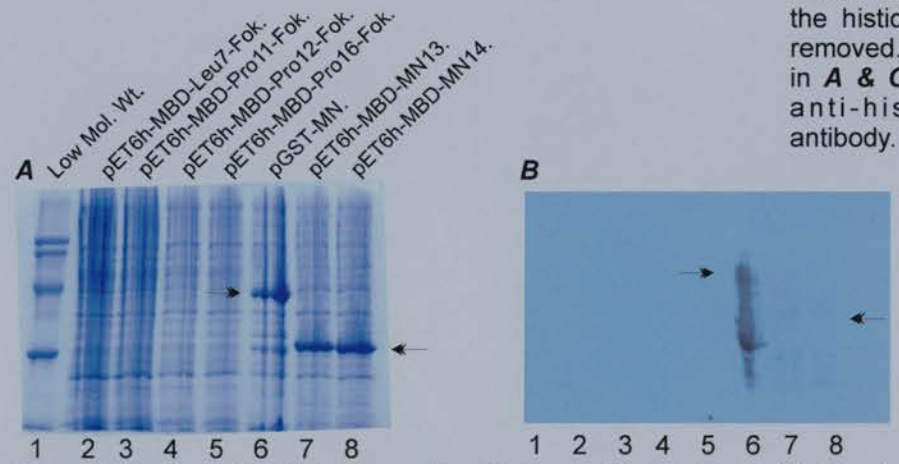


Figure 7.9 Construction of a methyl CpG specific nuclease. Verification of some attempts at analysis of fusion constructs in the pursuit of creating a methyl CpG specific nuclease. ***A,*** an 8 % SDS-PAGE gel of fusion constructs expressed in *E. coli* BL21(DE3)pLysS cells. pET6h, histidine fusion vector. pGST, GST fusion vector. MBD, methyl binding domain of MeCP2. MN, Micrococcal nuclease. Fok, *Fok*I cleavage domain. Pro and Leu refer to Proline and Leucine linker / hinge sequences for flexibility in the protein. Numbering, 7,11,12,13,14,16; refers to plasmid clone number. Western analysis with anti-histidine tag antibody weakly detects clones in lanes 7 & 8 in figure ***B,*** which is numbered to indicate labelling is the same as in ***A.*** Anti-GST antibody detects strongly the fusion protein in lane 6, with perhaps some free GST also. None of the new constructs show any expression with IPTG induction, nor reaction to antibodies.

It was thought sensible to sequence the vector to ensure it contained the full-length HIV-1 integrase sequence and was not producing a different protein product. The HIV-1 integrase protein was expressed from a pET15b vector (called pIN^{1-288} here) or pIN^{1-288}/ F185K/C280S (Jenkins *et al.,* 1996) and ligated at an *Nde*I-*Bam*H1 digested site. The HIV-1 integrase sequence was excised from pET15b by restriction digests with *Xba*I and *Bam*H1 and ligated into an *Xba*I and *Bam*H1 digested pBLUESCRIPT II KS (+/-) plasmid. The ligation reaction was transformed into *E. coli* XL1-blue cells, plated on LB plates with Ampicillin to 50 μg/ml and incubated at 37°C overnight. Single colonies were picked and grown up in 5 ml of LB/Amp. Minipreps were prepared of the plasmid DNA and the presence of the HIV-1 integrase DNA sequence tested for by restriction digest analysis (data not shown). The presence of the annealed sequence was verified. The pBS(HIV-1 integrase) plasmid was sent for sequencing. The results showed that the plasmid did indeed contain the HIV-1 integrase sequence.

Several possibilities could have been pursued with respect to optimising the Integrase assay. Some of these have been tested. It would have been advantageous to definitively prove whether the protein activity was low in my purified fractions. It may also have been the concentration and quality of components in the integration reaction. There may also have been problems with the PCR detection of integrants. All of these components would have taken some time to optimise. Unfortunately, it then appeared that we would not be able to investigate any methylation induced structural changes with an A-form detecting Ruthenium based probe. Requests for this chemical were denied. The pursuit of this technique was then questioned and this project unfortunately abandoned. Later, the existence of Benzonase proved to be a unique and informative way around the problem, and the pursuit of methylation induced structural studies was continued.

7.11 Creation of a CpG methylation specific restriction enzyme.

After leaving the Integrase Assay Project and alongside the Benzonase assay project, I attempted to create a restriction enzyme with the ability to cleave CpG methylated DNA specifically. No such restriction enzyme currently exists. This was a project

started by Dr. R. Meehan and continued in the laboratory by an honours student. The basic premise was to fuse the methyl-binding domain (MBD) of MeCP2 to the cleavage domain of *Fok*I restriction enzyme. Various constructs had already been made by others, containing either the GST- or HIS-tagged MBD of MeCP2 and/or the *Fok*I or Micrococcal (MNase) nuclease domain. It had been found that MBD-*Fok*I fusion constructs could not cleave CpG methylated DNA and a linker or hinge region may be required to enable both domains of the fusion protein to operate effectively. I was provided with a series of pET6h fusion proteins containing the MBD of MeCP2 fused to different types of linker region (proline or leucine rich), followed by the FokI nuclease domain. These had not all been tested conclusively. Plasmid constructs were tested for the correct size and orientation of the DNA insert. Several were then selected for analysis. Western blots were probed with anti-HIS or anti-GST antibodies. Unfortunately, only the only positive control clones reacted well with the antibodies. The results of one of these experiments is shown in figure 7.9. Due to the uncertainty of this project it was not pursued any further, but may provide an interesting side-project for other lab members. I would suggest two other approaches. (1) To manipulate *Sss*I methylase, which methylates a single CpG, and enable it to cleave CpG sites instead. This may be done by targeted mutation of the methyltransferase part of the enzyme or by targeted deletion coupled with the fusion of another cleavage domain, perhaps that of *Hha*I restriction enzyme. (2) Interestingly, because Benzonase has been shown to have heightened ability to cleave at methylated DNA to begin with, it would be interesting to both investigate whether Benzonase is some kind of methyl-binding protein and also to perhaps fuse the already sequence biased cutting domain with the recognition domain of MeCP2 or indeed, one of the other MBDs. No tool yet exists for the cleavage of every CpG in a sequence. The closest determination of methylated sites in a DNA sequence is performed by Bisulphite mapping. To this end, the development of a CpG specific nuclease would be highly advantageous in this field. Unfortunately it was not one I had time to pursue.

References

Alexeev, D.G., Lipanov, A.A., Skuratovskii, I.Y. Poly(dA).poly(dT) is a B-type double helix with a distinctively narrow minor groove. Nature. Feb 26-Mar 4;325(6107):821-3. (1987).

Akhtar, A., Zink, D., Becker, P.B. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell.* Feb;5(2):367-75 (2000).

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., Zoghbi, H.Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet.* Oct;23(2):185-8. (1999).

Antequera, F., Macleod, D., Bird, A.P. Specific protection of methylated CpGs in mammalian nuclei. *Cell* Aug 11;58(3):509-17 (1989).

Antequera, F. & Bird, A. CpG islands. *EXS* **64**, 169-185 (1993).

Aoki, A. & Takebayashi, S. Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 29(17), 3506-3512. (2001).

Arents, G., Moudrianakis, E.N. Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proc Natl Acad Sci U S A.* Nov 15;90(22):10489-93 (1993).

Arndt-Jovin, D.J., Udvardy, A., Garner, M.M., Ritter, S. & Jovin, T.M. Z-DNA binding and inhibition by GTP of Drosophila topoisomerase II. *Biochemistry* **32**, 4862-4872 (1993).

Arnott, S, Chandrasekaran, R, Birdsall, D.L., Leslie, A.G., Ratliff R.L. Left-handed DNA helices. *Nature* Feb 21;283(5749):743-5 (1980).

Avery, O.T., MacLeod, C.M., & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* **79,** 137-158 (1944).

Azorin, F. & Rich, A. Isolation of Z-DNA binding proteins from SV40 minichromosomes: evidence for binding to the viral control region. *Cell* **41**, 365-374 (1985).

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., Wyllie, A. Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* Dec 23;18(56):8044-7 (1999).

Ball, T.K., Saurugger, P.N., Benedik, M.J. The extracellular nuclease gene of Serratia marcescens and its secretion from Escherichia coli. *Gene.* 57(2-3):183-92 (1987).

Ballestar, E., Yusufzai, T.M., Wolffe, A.P. Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry.* Jun 20;39(24):7100-6. (2000).

Ballestar, E. & Wolffe, A.P. Methyl-CpG-binding proteins Targeting specific gene repression. Review. *Eur. J. Biochem.* **268**, 1-6 (2001).

Ban, C., Sundaralingam, M. Crystal structure of the self-complementary 5'-purine start decamer d(GCACGCGTGC) in the A-DNA conformation. II. *Biophys J.* Sep;71(3):1222-7 (1996).

Ban, C., Ramakrishnan, B., Sundaralingam, M. Crystal structure of the self-complementary 5'-purine start decamer d(GCGCGCGCGC) in the Z-DNA conformation. I. *Biophys J.* Sep;71(3):1215-21 (1996).

Barbour,V.M., Tufarelli, C., Sharpe, J.A., Smith, Z.E., Ayyub, H., Heinlein, C.A., Sloane-Stanley, J., Indrak, K., Wood, W.G., Higgs, D.R. alpha-thalassemia resulting from a negative chromosomal position effect. *Blood* **96**, (3) 800-807 (2000).

Behe, M., Felsenfeld, G. Effects of methylation on a synthetic polynucleotide: The B-Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. U S A,* 87 (3) 1619-1623 (1981).

Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene [see comments]. *Nature* **405**, 482-485 (2000).

Berger, I., Winston, W., Manoharan, R., Schwartz, T., Alfken, J., Kim, Y.G., Lowenhaupt, K., Herbert, A., Rich, A. Spectroscopic characterization of a DNA-binding domain, Z alpha, from the editing enzyme, dsRNA adenosine deaminase: evidence for left-handed Z-DNA in the Z alpha-DNA complex. *Biochemistry* **37** , 13313-13321 (1998).

Bestor, T.H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 9(16): 2395-402. Review. (2000).

Bestor, T.H. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **326**, 179-187 (1990).

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., Szyf, M. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature.* Feb 18;397(6720):579-83 (1999).

Bienvenu, T., Carrie, A., de Roux, N., Vinet, M.C., Jonveaux, P., Couvert, P., Villard Arzimanoglou, A., Beldjord, C., Fontes, M., Tardieu, M., Chelly, J. MECP2 mutations account for most cases of typical forms of Rett syndrome. *Hum Mol Genet.* May 22;9(9):1377-84 (2000).

Bingman, C., Li, X., Zon, G., Sundaralingam, M. Crystal and molecular structure of d(GTGCGCAC): investigation of the effects of base sequence on the conformation of octamer duplexes. *Biochemistry.* Dec 29;31(51):12803-12 (1992a).

Bingman, C., Jain, S., Zon, G., Sundaralingam, M. Crystal and molecular structure of the alternating dodecamer d(GCGTACGTACGC) in the A-DNA form: comparison with the isomorphous non-alternating dodecamer d(CCGTACGTACGG). *Nucleic Acids Res.* Dec 25;20(24):6637-47 (1992b).

Bird, A.P. DNA methylation patterns and epigenetic memory. *Genes Dev.* 16(1): 6-21. Review. (2002).

Bird, A.P., Taggart, M.H., Nicholls, R.D. & Higgs, D.R. Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J.* **6**, 999-1004 (1987).

Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., Renkawitz, R. The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. *J Biol Chem.* Nov 10;275(45):34963-7 (2000).

Bourc'his, D., Le Bourhis, D., Patin, D., Niveleau, A., Comizzoli, P., Renard, J.P., Viegas-Pequignot, E. Delayed and Incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. *Current Biology* 11, (19) 1542-1546 (2001).

Boyes, J., Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* Mar 22;64(6):1123-34 (1991).

Boyes, J., Bird, A. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* Jan;11(1):327-33 (1992).

Brennan, R.G., Westhof, E., Sundarlingam, M. Structure of a Z-DNA with two different backbone chain conformations. Stabilization of the decadeoxyoligonucleotide d(CGTACGTACG) by [Co(NH3)6]3+ binding to the guanine. *J Biomol Struct Dyn.* Feb;3(4):649-65 (1986).

Brenton, J.D., Drewell, R.A., Viville, S., Hilton, K.J., Barton, S.C., Ainscough, J.F., Surani, M.A. A silencer element identified in Drosophila is required for imprinting of H19 reporter transgenes in mice. *Proc Natl Acad Sci U S A.* Aug 3;96(16):9242-7 (1999).

Brown, B.A., Lowenhaupt, K., Wilbert, C.M., Hanlon, E.B. & Rich, A. The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA [In Process Citation]. *Proc. Natl. Acad. Sci. U. S. A* **97**, 13532-13536 (2000).

Brownell, J.E. & Allis, C.D. Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* 6(2): 176-84. Review. (1996).

Brukner, I., Jurukovski, V., Savic, A. Sequence dependent structural variations of DNA as revealed by DNaseI. *Nucleic Acids Res.* Feb 25; 18(4):891-4 (1990).

Bushman, F.D., Fujiwara, T. & Craigie, R. Retroviral DNA integration directed by HIV integration protein in vitro. *Science* **249**, 1555-1558 (1990).

Butkus, V., Klimasauskas, S., Petrauskiene, L., Maneliene, Z., Janulaitis, A., Minchenkova, L.E., Schyolkina, A.K. Synthesis and physical characterization of DNA fragments containing N4-methylcytosine and 5-methylcytosine. *Nucleic Acids Res.* Oct 26;15(20):8467-78 (1987).

Buyse, I.M., Fang, P., Hoon, K.T., Amir, R.E., Zoghbi, H.Y., Roa, B.B. Diagnostic testing for Rett syndrome by DHPLC and direct sequencing analysis of the MECP2 gene: identification of several novel mutations and polymorphisms. *Am J Hum Genet.* Dec;67(6):1428-36 (2000).

Casasnovas, J.M. & Azorin, F. Supercoiled induced transition to the Z-DNA conformation affects the ability of a d(CG/GC)12 sequence to be organized into nucleosome-cores . *Nucleic Acids Res.* **15**, 8899-8918 (1987).

Casasnovas, J.M., Ellison, M.J., Rodriguez-Campos, A., Martinez-Balbas, A. & Azorin, F. In vivo assessment of the Z-DNA-forming potential of d(CA.GT)n and d(CG.GC)n sequences cloned into SV40 minichromosomes. *J. Mol. Biol.* **208**, 537-549 (1989).

Chandler, S.P., Guschin, D., Landsberger, N. & Wolffe, A.P. The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry* **38**, 7008-7018 (1999).

Cheadle, J.P., Gill, H., Fleming, N., Maynard, J., Kerr, A., Leonard, H., Krawczak, M., Cooper, D.N., Lynch, S., Thomas, N., Hughes, H., Hulten, M., Ravine, D., Sampson J.R., Clarke, A. Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. *Hum Mol Genet.* Apr 12;9(7):1119-29 (2000).

Chen, R.Z., Akbarian, S., Tudor, M., Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genet.* Mar;27(3):327-31 (2001).

Cheng, X., Roberts, R.J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* 29(18), 3784-3795. (2001).

Chevrier, B., Dock, A.C., Hartmann, B., Leng, M., Moras, D., Thuong, M.T., Westhof, E. Solvation of the left-handed hexamer d(5BrC-G-5BrC-G-5 BrC-G) in crystals grown at two temperatures. *J. Mol. Biol.* Apr 20;188(4):707-19 (1986).

Chow, S.A. *In vitro* Assays for activities of retroviral integrase., in *Methods in Enzymology* 12, 306-317.review. (1997).

Christen, T., Bischoff, M., Hobi, R. & Kuenzle, C.C. High mobility group proteins 1 and 2 bind preferentially to brominated poly(dG-dC).poly(dG-dC) in the Z-DNA conformation but not to other types of Z-DNA. *FEBS Lett.* **267**, 139-141 (1990).

Clark, G.R., Brown, D.G., Sanderson, M.R., Chwalinski, T., Neidle, S., Veal, J.M., Jones, R.L., Wilson, W.D., Zon, G., Garman, E. Crystal and solution structures of the oligonucleotide d(ATGCGCAT)2: a combined X-ray and NMR study. *Nucleic Acids Res.* Sep 25;18(18):5521-8 (1990).

Coll, M., Fita, I., Lloveras, J., Subirana, J.A., Bardella, F., Huynh-Dinh, T., Igolen, J. Structure of d(CACGTG), a Z-DNA hexamer containing AT base pairs. *Nucleic Acids Res.* Sep 12;16(17):8695-705 (1988).

Collins, M., Myers, R.M. Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis. *J Mol Biol.* Dec 20;198(4):737-44 (1987).

Conner, B.N., Takano, T., Tanaka, S., Itakura, K., Dickerson, R.E. The molecular structure of d(CpCpGpG), a fragment of right-handed double helical A-DNA. *Nature* 295; 294-299 (1982).

Crawford, J.L., Kolpak, F.J., Wang, A.H., Quigley, G.J., van Boom, J.H., van der Marel G., Rich, A. The tetramer d(CpGpCpG) crystallizes as a left-handed double helix. *Proc Natl Acad Sci U S A.* Jul;77(7):4016-20 (1980).

Crick, F.H. DNA: test of structure? *Science* **167**, 1694 (1970).

Crick, F.H., Wang,J.C. & Bauer,W.R. Is DNA really a double helix? *J. Mol. Biol.* **129**, 449-457 (1979).

Cross, S.H., Meehan, R.R., Nan, X., Bird, A. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet.* Jul;16(3):256-9 (1997).

Crothers, D.M., Haran, T.E., Nadeau, J.G. Intrinsically bent DNA. *J. Biol. Chem.* 265 (13); 7093-7096 (1990).

Csordas, A. On the biological role of histone acetylation. *Biochem J.* 265(1): 23-38. Review. (1990).

Cuadrado, M., Sacristan, M., Antequera, F. Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep.* 2001 Jul;2(7):586-92.

Daniel, J.M., Reynolds, A.B. The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol.* May;19(5):3614-23 (1999).

Davey, C., Pennings, S., Meersseman, G., Wess, T.J., Allan, J. Periodicity of strong nucleosome positioning sites around the chicken adult beta-globin gene may encode regularly spaced chromatin. *Proc. Natl. Acad. Sci U S A* Nov 21;92(24):11210-4 (1995).

Davey, C., Pennings, S., Allan, J. CpG methylation remodels chromatin structure in vitro. *J.Mol.Biol.* 267; 276-288 (1997).

Dean, W. Conservation of methylation reprogramming in mammalian development: Aberrant reprogramming in cloned embryos. *Proc. Natl Acad. of Sci U S A* 98(24), 13734-13738. (2001).

Di Mauro, E., Verdone, L., Chiappini, B. & Caserta, M. In Vivo Changes of Nucleosome Positioning in the Pretranscription State. *J. Biol. Chem.* **277**, 7002-7009 (2002).

Dickerson R.E., Ng, H.L. DNA structure from A to B. *Proc Natl Acad Sci U S A.* Jun 19;98(13):6986-8. Review. (2001).

Diekmann, S. DNA methylation can enhance or induce DNA curvature. *EMBO J.* 6(13): 4213-7. (1987).

Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J.S., Gough, G.R., Gilham, P.T., Moras, D. Crystallographic structure of an RNA helix [U(U-A)6A]2. *J. Mol. Biol.* Oct 5;209(3): 459-74 (1989).

Drew, H.R. & Travers, A.A. DNA structural variations in the *E. coli* tyrT promoter. *Cell.* 37(2): 491-502 (1984).

Drew,H., Takano,T., Tanaka,S., Itakura,K. & Dickerson,R.E. High-salt d(CpGpCpG), a left-handed Z-DNA double helix. *Nature* **286**, 567-573 (1980).

Drewell, R.A., Brenton J.D., Ainscough, J.F., Barton, S.C., Hilton, K.J., Arney, K.L., Dandolo, L., Surani, M.A. Deletion of a silencer element disrupts H19 imprinting independently of a DNA methylation epigenetic switch. *Development* **127**, 3419-3428 (2000).

Drewell, R.A., Arney, K.L., Arima, T., Barton, S.C., Brenton, J.D., Surani, M.A. Novel conserved elements upstream of the H19 gene are transcribed and act as mesodermal enhancers. *Development.* Mar;129(5):1205-13. (2002).

Ebralidse, K.K., Grachev, S.A., Mirzabekov, A.D. A highly basic histone H4 domain bound to the sharply bent region of nucleosomal DNA. *Nature.* Jan 28; 331 (6154): 365-7.(1988).

Ehrlich, M., Ehrlich, K., Mayo, J.A.Unusual properties of the DNA from Xanthomonas phage XP-12 in which 5-methylcytosine completely replaces cytosine. *Biochim Biophys Acta.* Jun 16;395(2):109-19.(1975).

Ehrlich, M., Zhang, X.Y. & Wang, R.Y. Human DNA methylation: methylated DNA-binding protein, differentiation and cancer. *Prog. Clin. Biol. Res.* **198**, 255-269 (1985).

Eisenstein, M., Shakked, Z. Hydration pattens and intermolecular interactions in A-DNA crystal structures. Implications for DNA recognition. *J.Mol.Biol.*, 248; 662-678 (1995).

El Hassan, M.A., Calladine, C.R. Two distinct modes of protein-induced bending in DNA. *J Mol Biol.* Sep 18;282(2):331-43. Review. (1998).

Esteller M. & Herman, J.G. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J.Pathol.* 196(1), 1-7. (2002).

Fabriciova, G., Miskovsky, P., Jancura, D., Lisy, V.Characterization of low-salt and high-salt conformation of poly(dI-dC) by hydrogen-deuterium exchange kinetics: a classical Raman spectroscopy study. *J Biomol Struct Dyn.* Oct;16(2):281-8. (1998).

Fedoroff, N. Transposons and genome evolution in plants. *P.N.A.S.* June 20, 97(13), 7002-7. Review. (2002).

Feigon, J., Wang, A.H., van der Marel, G.A., van Boom, J.H. & Rich, A. Z-DNA forms without an alternating purine-pyrimidine sequence in solution. *Science* **230**, 82-84 (1985).

Feng, Q., Zhang, Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.* Apr 1;15(7):827-32 (2001).

Ferguson-Smith, A.C. & Surani, M.A. Imprinting and the Epigenetic Asymmetry Between Parental Genomes. *Science* **293**, 1086-1089 (2001).

Finerty, P.J. Jr., Bass, B.L. A Xenopus zinc finger protein that specifically binds dsRNA and RNA-DNA hybrids. *J. Mol. Biol.* Aug 15; 271(2):195-208 (1997).

Finerty, P.J. Jr, Bass, B.L. Subsets of the zinc finger motifs in dsRBP-ZFa can bind double-stranded RNA. *Biochemistry*. Mar 30;38(13):4001-7 (1999).

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. Potent and specific genetic interferance by double-stranded RNA in Caenorhabditis elegans. *Nature* Feb 19; 391 (6669): 806-11 (1998).

Fischel-Ghodsian, N., Nicholls, R.D. & Higgs,D.R. Long range genome structure around the human alpha-globin complex analysed by PFGE. *Nucleic Acids Res.* **15**, 6197-6207 (1987).

Foloppe, N., MacKerell, AD Jr. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys J.*, 76(6): 3206-18. (1999).

Fox, K.R, Waring, M.J. DNA structural variations produced by actinomycin and distamycin as revealed by DNAase I footprinting. *Nucleic Acids Res.* Dec 21;12(24):9271-85 (1984).

Fox, K.R. The effect of HhaI methylation on DNA local structure. *Biochem. J.* **234**, 213-216 (1986).

Franke, I., Meiss, G., Blecher, D., Gimadutdinow, O., Urbanke, C., Pingoud, A. Genetic engineering, production and characterisation of monomeric variants of the dimeric Serratia marcescens endonuclease. *FEBS Lett.* Apr 3;425(3):517-22 (1998).

Franklin, R.E. & Gosling, R.G. Evidence for a 2-chain helix in the crystalline structure of Sodium deoxyribonucleate. *Nature* 172: 156 (1953).

Free, A., Wakefield, R.I., Smith, B.O., Dryden, D.T., Barlow, P.N., Bird, A.P. DNA recognition by the methyl-CpG binding domain of MeCP2. *J Biol Chem.* Feb 2;276(5):3353-60 (2001).

Friedhoff, P., Gimadutdinow, O., Pingoud, A. Identification of catalytically relevant amino acids of the extracellular Serratia marcescens endonuclease by alignment-guided mutagenesis. *Nucleic Acids Res.* Aug 25;22(16):3280-7 (1994).

Fujii, S., Wang, A.H., van der, M.G., van Boom, J.H. & Rich, A. Molecular structure of (m5 dC-dG)3: the role of the methyl group on 5- methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.* **10**, 7879-7892 (1982).

Fujii, S., Wang, A.H., Quigley, G.J., Westerink, H., Van der Marel, G., Van Boom, J.H. The octamers d(CGCGCGCG) and d(CGCATGCG) both crystallize as Z-DNA in the same hexagonal lattice. *Biopolymers.* Jan;24(1):243-5 (1985).

Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., Nakao, M. Methylation-mediated transcriptional silencing in euchromatin by methyl- CpG binding protein MBD1 isoforms. *Mol. Cell Biol.* **19**, 6415-6426 (1999).

Gagna, C.E., Chen, J.H., Kuo, H.R. & Lambert,W.C. Binding properties of bovine ocular lens zeta-crystallin to right- handed B-DNA, left-handed Z-DNA, and single-stranded DNA. *Cell Biol. Int.* **22**, 217-225 (1998).

Garner, M.M. & Felsenfeld, G. Effect of Z-DNA on nucleosome placement. *J. Mol. Biol.* **196**, 581-590 (1987).

Ginder GD, Singal R, Little J.A., Dempsey, N., Ferris, R., Wang, S.Z. Silencing and activation of embryonic globin gene expression. *Ann N Y Acad Sci.* Jun 30;850:70-9. Review (1998).

Glikin G.C., Jovin, T.M., Arndt-Jovin, D.J. Interactions of Drosophila DNA topoisomerase II with left-handed Z-DNA in supercoiled minicircles. *Nucleic Acids Res.* Dec;19(25):7139-44 (1991).

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., Mello, C.C. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell,* 106(1): 23-34 (2001).

Gross, M.M., Janiaud, P. & Stanbridge, E.J. Effect of retroviral integration on control of expression of a tumor marker in HeLa cells. *Mol. Carcinog.* **8**, 89-95 (1993).

Groudine, M. Chromatin structure and developmental expression of the human alpha-globin cluster. *Mol Cell Biol.* Apr;6(4):1108-16 (1986).

Gut, S.H., Bischoff, M., Hobi, R. & Kuenzle, C.C. Z-DNA-binding proteins from bull testis. *Nucleic Acids Res.* **15**, 9691-9705 (1987).

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet.* 2001 Mar;27(3):322-6.

Hagerman, P.J. Pyrimidine 5-methyl groups influence the magnitude of DNA curvature. *Biochemistry* **29**, 1980-1983 (1990a).

Hagerman, P.J. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59**, 755-781 (1990b).

Haigh, L.S., Owens, B.B., Hellewell, O.S. & Ingram, V.M. DNA methylation in chicken alpha-globin gene expression. *Proc. Natl. Acad. Sci. U. S. A* **79**, 5332-5336 (1982).

Hampson, K., Woods, C.G., Latif, F., Webb, T. Mutations in the MECP2 gene in a cohort of girls with Rett syndrome. *J Med Genet.* Aug;37(8):610-2 (2000).

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., Tilghman, S.M. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature.* May 25; 405(6785):486-9 (2000).

Heinemann, U., Alings, C., Hahn, M. Crystallographic studies of DNA helix structure. *Biophys Chem.* May;50(1-2):157-67 (1994).

Heinemann, U., Lauble, H., Frank, R., Blocker, H. Crystal structure analysis of an A-DNA fragment at 1.8 A resolution: d(GCCCGGGC). *Nucleic Acids Res.* Nov 25;15(22):9531-50 (1987).

Heinemann, U., Rudolph, L.N., Alings, C., Morr, M., Heikens, W., Frank, R., Blocker, H. Effect of a single 3'-methylene phosphonate linkage on the conformation of an A-DNA octamer double helix. *Nucleic Acids Res.* Feb 11;19(3):427-33 (1991).

Hendrich, B., Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol.* Nov;18(11):6538-47 (1998).

Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S., Bird, A. Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mamm. Genome* **10**, 906-912 (1999).

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.*Genes Dev.* Mar 15;15(6):710-23 (2001).

Herbert, A., Lowenhaupt, K., Spitzner, J. & Rich, A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl. Acad. Sci. U. S. A* **92**, 7550-7554 (1995a).

Herbert, A., Lowenhaupt, K., Spitzner, J. & Rich, A. Double-stranded RNA adenosine deaminase binds Z-DNA in vitro. *Nucleic Acids Symp. Ser.* 16-19 (1995b).

Herbert, A., Rich, A. The biology of left-handed Z-DNA. *J Biol Chem.* May 17;271(20):11595-8. Review. (1996).

Herbert, A., Alfken, J., Kim, Y.G., Mian, I.S., Nishikura, K., Rich, A. A Z-DNA binding domain present in the human editing enzyme, double- stranded RNA adenosine deaminase. *Proc. Natl. Acad. Sci. U. S. A* **94**, 8421-8426 (1997).

Herbert, A., Schade, M., Lowenhaupt, K., Alfken, J., Schwartz, T., Shlyakhtenko, L.S., Lyubchenko, Y.L., Rich, A. The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences. *Nucleic Acids Res.* **26**, (15) 3486-3493 (1998).

Herbert, A. & Rich, A. Left-handed Z-DNA: structure and function. *Genetica* **106**, 37-47 (1999).

Herbert, A. & Rich, A. The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *PNAS* **98**, 12132-12137 (2001).

Higgs, D.R., Vickers, M.A., Wilkie, A.O., Pretorius, I.M., Jarman, A.P., Weatherall, D.J. A review of the molecular genetics of the human alpha-globin gene cluster. *Blood.* Apr;73(5):1081-104. Review. (1989).

Hill, R.J. Z-DNA: A prodrome for the 1990s. *J. Cell Science*, 99, 675-680 (1991).

Hill, R.J. & Stollar, B.D. Dependence of Z-DNA antibody binding to polytene chromosomes on acid fixation and DNA torsional strain. *Nature* **305**, 338-340 (1983).

Ho, P.S., Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H. G.T wobble base-pairing in Z-DNA at 1.0 A atomic resolution: the crystal structure of d(CGCGTG). *EMBO J.* Dec 16;4(13A):3617-23 (1985).

Hobert, O., Ruvkun, G. A common theme for LIM homeobox gene function across phylogeny? *Biol Bull.* Dec; 195(3):377-80. Review. (1998).

Hodges-Garcia, Y. & Hagerman, P.J. Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry* **31**, 7595-7599 (1992).

Hodges-Garcia, Y. & Hagerman, P.J. Investigation of the Influence of Cytosine Methylation on DNA Flexibility. *J. Biol. Chem.* **270**, 197-201 (1995).

Hogan, M.E., Roberson, M.W., Austin, R.H. DNA flexibility variation may dominate DNase I cleavage. *Proc Natl Acad Sci U S A.* Dec;86(23):9273-7 (1989).

Hogan, M.E., Rooney, T.F., Austin, R.H. Evidence for kinks in DNA folding in the nucleosome. *Nature.* Aug 6-12;328(6130):554-7 (1987).

Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science.* Jan 24; 187 (4123): 226-232. (1975).

Hough, R.F., Bass, B.L. Analysis of Xenopus dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases. *RNA.* Apr; 3(4): 356-70 (1997).

Hovatter, K.R., Martinson, H.G. Ribonucleotide-induced helical alteration in DNA prevents nucleosome formation. *Proc Natl Acad Sci U S A.* Mar;84(5):1162-6 (1987).

Hsieh, C.L. Dynamics of DNA methylation pattern. *Curr. Opin. Genet. Dev.* 10(2): 86-91. Review. (2000).

Hsieh, C.L. Stability of patch methylation and its impact in regions of transcriptional initiation and elongation. *Mol. Cell Biol.* **17**, 5897-5904 (1997).

Hua, N., van der Marel, G.A., van Boom, J.H. & Feigon, J. Non-contiguous regions of Z-DNA in a DNA dodecamer. *Nucleic Acids Res.* **17**, 7923-7944 (1989).

Huang, N., vom Baur, E., Garnier, J.M., Lerouge, T., Vonesch, J.L., Lutz, Y.,Chambon P., Losson, R. Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.* Jun 15;17(12):3398-412 (1998).

Huppke, P., Laccone, F., Kramer, N., Engel, W., Hanefield, F. Rett syndrome: analysis of MeCP2 and clinical characterization of 31 patients. *Hum. Mol. Genet.* May 22; 9(9):1369-75 (2000).

Itakura, K., Katagiri, N., Narang, S.A., Bahl, C.P., Marians, K.J., Wu, R. Chemical synthesis and sequence studies of deoxyribooligonucleotides which constitute the

duplex sequence of the lactose operator of Escherichia coli. *J Biol Chem.* Jun 25;250(12):4592-600 (1975).

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovski, G., Dausman, J., Lee, P., Wilson, C., Lander, E., Jaenisch, R. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* 27(1):31-9. (2001).

Jain, S., Zon, G., Sundaralingam, M. The potentially Z-DNA-forming sequence d(GTGTACAC) crystallizes as A-DNA. *J Mol Biol.* Sep 5;197(1):141-5 (1987).

Jamin, N., James, T.L., Zon, G. Two-dimensional nuclear Overhauser enhancement investigation of the solution structure and dynamics of the DNA octamer [d(GGTATACC)]2. *Eur J Biochem.* Oct 1;152(1):157-66 (1985).

Jenkins, T., Engelman, A., Ghirlando, R., Craigie, R. A soluble active mutant of HIV-1 integrase. *J.Biol.Chem.* 271 (13), 7712-7718 (1996).

Jenuwein, T. & Allis, C.D. Translating the Histone Code. *Science* **293**, 1074-1080 (2001).

Jones, P.A., Veenstra, G., Wade, P., Vermaak, D., Kass, S., Landsberger, N., Strouboulis, J., Wolffe, A. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, 19; 187-191 (1998).

Jones, P.A., Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* Jun; 3(6):415-28. Review. (2002).

Kanduri, C., Holmgren, C., Pilartz, M., Franklin, G., Kanduri, M., Liu, L., Ginjala, V., Ulleras, E., Mattsson, R., Ohlsson, R. The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. *Curr. Biol.* Apr 20;10(8): 449-57 (2000a).

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C.F., Wolffe, A., Ohlsson, R., Lobanenkov, V.V. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol.* Jul 13;10(14):853-6 (2000b).

Kass, S.U., Landsberger, N., Wolffe, A.P.DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol.* Mar 1;7(3):157-65 (1997).

Kellum, R. & Schedl, P. A position effect assay for boundaries of higher order chromosomal domains. *Cell,* 64(5): 941-50. (1991).

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Philips, D.C., Shore, V.C. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature,* 185, 422-427 (1960).

Kim, S.J., Cook, E.H. Jr. Novel de novo nonsense mutation of MeCP2 in a patient with Rett syndrome. *Hum. Mutat.* Apr;15(4):382-3 (2000).

Kitamura, Y., Ha Lee, Y.M., Coffin, J.M. Nonrandom integration of retroviral DNA in vitro: Effect of CpG methylation. *Proc. Natl. Acad. Sci. U S A.*, 89; 5532-5536 (1992).

Klysik, J., Stirdivant, S.M., Singleton, C.K., Zacharias, W., Wells, R.D. Effects of 5 cytosine methylation on the B-Z transition in DNA restriction fragments and recombinant plasmids. *J. Mol. Biol.* Jul 25; 168(1):51-71 (1983).

Kokura, K., Kaul, S.C., Wadhwa, R., Nomura, T., Kahn, M.M., Shinagawa, T., Yasukawa, T., Colmenares, C., Ishii, S. The Ski protein family is required for

MeCP2-mediated transcriptional repression. *J. Biol. Chem.* Sep 7;276(36):34115-21 (2001).

Koo, H.S., Crothers, D.M. Chemical determinants of DNA bending at adenine-thymine tracts. *Biochemistry.* Jun 16;26(12):3745-8 (1987).

Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science.* 184: 868-871. (1974).

Kornberg, R.D. Structure of chromatin. *Ann. Rev. Biochem.* 46: 931-954 (1977).

Kress, C., Thomassin, H., Grange, T. Local DNA demethylation in vertebrates: how could it be performed and targeted? *FEBBS Lett.* 494(3): 135-40. Review. (2001).

Kumar, G.S., Debnath, D., Maiti, M. Conformational aspects of poly(dI-dC).poly(dI-dC) and poly(dG-dC).poly(dG-dC) on binding of the alkaloid, berberine chloride. *Anticancer Drug Des.* Aug;7(4):305-14 (1992).

Lafer, E.M., Valle, R.P., Moller, A. Nordheim, A., Schur, P.H., Rich, A., Stollar, B.D. Z-DNA-specific antibodies in human systemic lupus erythematosus. *J Clin Invest.* Feb;71(2):314-21 (1983).

Lancillotti, F., Lopez, M.C., Alonso, C. & Stollar, B.D. Locations of Z-DNA in polytene chromosomes. *J. Cell Biol.* **100**, 1759-1766 (1985).

Lancillotti, F., Lopez, M.C., Arias, P. & Alonso, C. Z-DNA in transcriptionally active chromosomes. *Proc. Natl. Acad. Sci. U. S. A* **84**, 1560-1564 (1987).

Langlois, d'Estaintot, B., Dautant, A., Courseille, C., Precigoux, G. Orthorhombic crystal structure of the A-DNA octamer d(GTACGTAC). Comparison with the tetragonal structure. *Eur J Biochem.* Apr 15;213(2):673-82 (1993).

Lefebvre, A., Mauffret, O., Hartmann, B., Lescot, E., Fermandjian, S. Structural behavior of the CpG step in two related oligonucleotides reflects its malleability in solution. *Biochemistry.* Sep 19;34(37):12019-28. (1995a)

Lefebvre, A., Mauffret, O., el Antri, S., Monnot, M., Lescot, E., Fermandjian, S. Sequence dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study. *Eur J Biochem.* Apr 15;229(2):445-54. (1995b).

Lewin, B. Genes VII. *Oxford University Press.* 2000.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy I., Jeppesen, P., Klein, F., Bird, A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905-914 (1992).

Lim, A.C., Barton, J.K. Rh(phen)2phi3+ as a shape-selective probe of triple helices. *Biochemistry.* Jun 23;37(25):9138-46 (1998).

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., Jacobsen, S.E. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science.* Jun 15;292(5524):2077-80 (2001).

Liu, L.F., Wang, J.C. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A.* Oct;84(20):7024-7 (1987).

Lorincz, M.C., Groudine, M. C(m)C(a/t)GG methylation: a new epigenetic mark in mammalian DNA? *Proc. Natl. Acad. Sci U S A.* 98(18): 10034-6. (2001).

Low, D.A., Weyand, N.J., Mahan, M.J. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect. Immun.* 69(12): 7197-204. Review. (2001).

Lunin, V.Y., Levdikov, V.M., Shlyapnikov, S.V., Blagova, E.V., Lunin, V.V., Wilson K.S., Mikhailov, A.M. Three-dimensional structure of Serratia marcescens nuclease at 1.7 A resolution and mechanism of its action. *FEBS Lett.* Jul 21;412(1):217-22 (1997).

Lustig, A.J. Mechanisms of silencing in Saccharomyces cerevisiae. *Curr. Opin. Genet. Dev.* 8(2): 233-9. (1998).

Ma, C., Sun, L. & Bloomfield, V.A. Condensation of plasmids enhanced by Z-DNA conformation of d(CG)n inserts. *Biochemistry* **34**, 3521-3528 (1995).

Mabuchi, H., Fujii, H., Calin, G., Alder, H., Negrini, M., Rassenti, L., Kipps, T.J., Bullrich, F., Croce, C.M. Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. *Cancer Res.* Apr 1;61(7): 2870-7 (2001).

Macleod, D., Clark, V.H., Bird, A. Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nat Genet.* Oct;23(2):139-40 (1999).

Malinina, L., Fernandez, L.G., Huynh-Dinh, T., Subirana, J.A. Structure of the d(CGCCCGCGGGCG) dodecamer: a kinked A-DNA molecule showing some B-DNA features. *J Mol Biol.* Jan 29;285(4):1679-90. (1999).

Martin, C.C., Laforest, L., Akimento, M.A., Ekker, M. A role for DNA methylation in gastrulation and somite patterning. *Dev. Biol.* Feb 15; 206(2):189-205 (1999).

Mattick, J.S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO J* 2(11), 986-991. (2001).

Matzke, M., Matzke, A.J.M. & Kooter, J.M. RNA: Guiding Gene Silencing . *Science* **293**, 1080-1083 (2001).

Mayer-Jung, C., Moras, D. & Timsit, Y. Hydration and recognition of methylated CpG steps in DNA. *EMBO J.* **17**, 2709-2718 (1998).

McClintock, B. The suppressor-mutator system of control of gene action in maize. *Carnegie Inst. of Wash. Year Book* 57: 415-429 (1958).

Meehan, R.R., Lewis, J.D., McKay,S., Kleiner,E.L. & Bird,A.P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499-507 (1989).

Meehan R., Lewis J., Jeppesen P., Bird A. Methylated DNA-binding proteins and chromatin structure (1993). *Chromosomes Today*, 11, 377-389 (1993).

Meehan, R. & Stancheva, I. DNA methylation and control of gene expression in vertebrate development. *Essays Biochem.* 37, 59-70. Review (2001).

Meehan, R., Lewis, J., Cross, S., Nan, X., Jeppesen, P., Bird, A. Transcriptional repression by methylation of CpG. *J. Cell Sci. Suppl* **16**, 9-14 (1992).

Meehan, R.R., Lewis, J.D. & Bird, A.P. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* **20**, 5085-5092 (1992).

Mei, H.Y., Barton, J.K. Tris(tetramethylphenanthroline)ruthenium(II): a chiral probe that cleaves A-DNA conformations. *Proc Natl Acad Sci U S A.* Mar;85(5):1339-43 (1988).

Meiss, G., Friedhoff, P., Hahn, M., Gimadutdinow, O. & Pingoud, A. Sequence preferences in cleavage of dsDNA and ssDNA by the extracellular Serratia marcescens endonuclease. *Biochemistry* **34**, 11979-11988 (1995).

Meiss, G., Franke, I., Gimadutdinow, O., Urbanke, C., Pingoud, A. Biochemical characterization of Anabaena sp. strain PCC 7120 non-specific nuclease NucA and its inhibitor NuiA. *Eur J Biochem.* Feb 1;251(3):924-34 (1998).

Meiss G, Gast FU, Pingoud AM. The DNA/RNA non-specific Serratia nuclease prefers double-stranded A-form nucleic acids as substrates. *J Mol Biol.* May 7; 288 (3):377-90 (1999).

Mette, MF., Aufsatz, W., van der Winden, J., Matzke, M.A., Matzke, A.J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* 19(19): 5194-201. (2000).

Michalowsky, L.A. & Jones, P.A. DNA methylation and differentiation. *Environ. Health Perspect.* **80**, 189-197 (1989).

Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J., Krause, K.L. 2.1 Å structure of Serratia endonuclease suggests a mechanism for binding to double-stranded DNA. *Nat Struct Biol.* Jul;1(7):461-8 (1994).

Miller, F.D., Dixon, G.H., Rattner, J.B. & Van de Sande, J.H. Assembly and characterization of nucleosomal cores on B- vs. Z-form DNA. *Biochemistry* **24**, 102-109 (1985).

Mohr, S.C., Sokolov, N.V., He, C.M., Setlow, P. Binding of small acid-soluble spore proteins from Bacillus subtilis changes the conformation of DNA from B to A. *Proc Natl Acad Sci U S A.* Jan 1;88(1):77-81 (1991).

Mooers, B.H.M., Schroth, G.P., Baxter, W.W. Alternaating and non-alternating dG-dC hexanucleotides cyrstallize as canonical A-DNA. *J. Mol. Biol.*, 249, 772-784 (1995).

Morel, J., Mourrain, P., Beclin, C. & Vaucheret, H. DNA methylation and chromatin structure affect transcriptional and post- transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* **10**, 1591-1594 (2000).

Muiznieks, I. & Doerfler, W. The topology of the promoter of RNA polymerase II- and III-transcribed genes is modified by the methylation of 5'-CG-3' dinucleotides. *Nucleic Acids Res.* **22**, 2568-2575 (1994).

Muller, H.P. & Varmus, H.E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13**, 4704-4714 (1994).

Mura, C.V. & Stollar, B.D. Interactions of H1 and H5 histones with polynucleotides of B- and Z-DNA conformations. *Biochemistry* **23**, 6147-6152 (1984).

Murchie, A.I., Lilley, D.M. Base methylation and local DNA helix stability. Effect on the kinetics of cruciform extrusion. *J.Mol.Biol.*, 205, 593-602 (1989).

Murchie, A.I, Lilley, D.M.The mechanism of cruciform formation in supercoiled DNA: initial opening of central basepairs in salt-dependent extrusion. *Nucleic Acids Res.* Dec 10;15(23):9641-54 (1987).

Nakao, M. Epigenetics: interaction of DNA methyation and chromatin. *Gene* 278, 25-31., (2001).

Nan, X., Meehan, R.R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21**, 4886-4892 (1993).

Nan, X., Tate, P., Li, E., Bird, A. DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol.* Jan;16(1):414-21 (1996).

Nan, X., Campoy, F.J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471-481 (1997).

Nan, X., Hg, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A. Transcriptional repression by the methyl-CpG binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393(6683): 386-9. (1998).

Nestle, M., Roberts, W.K. An extracellular nuclease from Serratia marcescens. I. Purification and some properties of the enzyme. *J Biol Chem.* Oct 10;244(19):5213-8 (1969a).

Nestle, M., Roberts, W.K. An extracellular nuclease from Serratia marcescens. II. Specificity of the enzyme. *J Biol Chem.* Oct 10;244(19):5219-25 (1969b).

Ng, H.H. & Bird, A. DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.* **9**, 158-163 (1999).

Ng, H.H., Jeppesen, P. & Bird, A. Active repression of methylated genes by the chromosomal protein MBD1. *Mol. Cell Biol.* **20**, 1394-1406 (2000).

Nightingale, K., Wolffe, A.P. Methylation at CpG sequences does not influence histone H1 binding to a nucleosome including a Xenopus borealis 5 S rRNA gene. *J Biol Chem.* Mar 3;270(9):4197-200 (1995).

Nordheim, A., Pardue, M.L., Lafer, E.M., Moller, A., Stollar, B.D., Rich, A. Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes. *Nature* **294**, (5840) 417-422 (1981).

Obata, K., Matsuishi, T., Yamashita, Y., Fukuda, T., Kuwajima, K., Honuchi, I., Nagamitsu, S., Iwanaga, R., Kimura, A., Omori, I., Endo, S., Mori, K., Kondo, I. Mutation analysis of the methyl-CpG binding protein 2 gene (MeCP2) in patients with Rett syndrome. *J. Med. Genet.* Aug;37(8):608-10 (2000).

Ohishi, H., Kunisawa, S., van der Marel, G., van Boom, J.H., Rich, A., Wang, A.H., Tomita, K., Hakoshima, T. Interaction between the left-handed Z-DNA and polyamine. The crystal structure of the d(CG)3 and N-(2-aminoethyl)-1,4-diamino-butane complex. *FEBS Lett.* **284**, (2):238-244 (1991).

Ohishi, H., Terasoma, N., Nakanishi, I., van der Marel, G., van Boom, J.H., Rich, A., Wang, A.H., Hakoshima, T., Tomita, K. Interaction between left-handed Z-DNA and polyamine - 3. The crystal structure of the d(CG)3 and thermospermine complex. *FEBS Lett.* **398**, (2-3):291-296 (1996).

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., Shirakawa, M. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* 18;105(4):487-97 (2001).

Ohki, I., Shimotake, N., Fujita, N., Nakao, M. & Shirakawa, M. Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J.* **18**, 6653-6661 (1999).

Ohyama, T. Intrinsic DNA bends: an organizer of local chromatin structure for transcription. *Bioessays* Aug;23(8):708-15. Review (2001).

Olson, W.K. The spatial configuration of ordered polynucleotide chains. II. The poly(rA) helix. *Nucleic Acids Res.* Nov;2(11):2055-68 (1975).

Pannell, D. & Ellis, J. Silencing of gene expression: implications for design of retrovirus vectors. *Rev. Med. Virol.* 11(4): 205-17. Review. (2001).

Pardue, M.L., Nordheim, A., Lafer, E.M., Stollar, B.D. & Rich, A. Z-DNA and the polytene chromosome. *Cold Spring Harb. Symp. Quant. Biol.* **47 Pt 1**, 171-176 (1983).

Park,Y. & Kuroda,M.I. Epigenetic Aspects of X-Chromosome Dosage Compensation. *Science* **293**, 1083-1085 (2001).

Pohl, F.M., Jovin, T.M. Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC). *J Mol Biol.* Jun 28;67(3):375-96 (1972).

Pondel, M.D., Murphy, S., Pearson, L., Craddock, C., Proudfoot, N.J. Sp1 functions in a chromatin-dependent manner to augment human alpha-globin promoter activity. *Proc. Natl. Acad. Sci. U S A* Aug 1;92(16):7237-41 (1995).

Prive, G.G., Yanagi, K., Dickerson, R.E. Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and CC-A-G-G-C-C-T-G-G. *J Mol Biol.* Jan 5;217(1):177-99 (1991).

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., Prokhortchouk, E. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* Jul 1;15(13):1613-8 (2001).

Pruss, D., Bushman, F.D., Wolffe, A.P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. U S A*, 91; 5913-5917 (1994a).

Pruss, D., Reeves, R., Bushman, F.D., Wolffe, A.P. The influence of DNA and nucleosome structure on integration events directed by HIV-integrase. *J. Biol. Chem.* 269, 25031-25041 (1994b).

Rea, S., Eisenhaber, F., O' Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., Jenuwein, T. Regulation of chromatin structure by site specific histone H3 methyltransferases. *Nature* Aug 10;406(6796):593-9 (2000).

Reik, W., Dean, W. & Walter, J. Epigenetic Reprogramming in Mammalian Development. *Science* **293**, 1089-1093 (2001).

Rein, T., Forster, R., Krause, A., Winnacker, E.L. & Zorbas, H. Organization of the alpha-globin promoter and possible role of nuclear factor I in an alpha-globin-inducible and a noninducible cell line. *J. Biol. Chem.* **270**, 19643-19650 (1995).

Riccio, A., Aaltonen, L.A., Godwin, A.K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D.E., Ruggeri, B.A, Klein-Szanto, A.J., Testa, J.R., Neri, G., Bellacosa, A. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite *Nat Genet.* Nov;23(3):266-8 (1999).

Rich, A., Speculation on the biological roles of left-handed Z-DNA. In*: "DNA damage. Effects on DNA structure and protein recognition"*. eds: Wallace S., Van Houten B., Wah Kow Y., 726, 1-17 (1994).

Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9-25 (1975).

Rohner, K.J., Hobi, R. & Kuenzle, C.C. Z-DNA-binding proteins. Identification critically depends on the proper choice of ligands. *J. Biol. Chem.* **265**, 19112-19115 (1990).

Runkel, L. & Nordheim,A. Chemical footprinting of the interaction between left-handed Z-DNA and anti-Z-DNA antibodies by diethylpyrocarbonate carbethoxylation. *J. Mol. Biol.* **189**, 487-501 (1986).

Sadasivan, C., Gautham, N. Sequence-dependent microheterogeneity of Z-DNA: the crystal and molecular structures of d(CACGCG).d(CGCGTG) and d(CGCACG).d(CGTGCG). *J Mol Biol.* May 19;248(5):918-30 (1995).

Sambrook, J., Fritsch, E.F., Maniatis, T. Molecular Cloning, A laboratory manual. 3rd Edition. *CSHLP* (1997).

Santoro, R., Grummt, I. Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol. Cell.* 8(3): 719-25. (2001).

Schade, M., Turner, C.J., Kuhne, R., Schmeider, P., Lowenhaupt, K., Herbert, A., Rich, A., Oschkinat, H. The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl. Acad. Sci. U. S. A* **96**, 12465-12470 (1999a).

Schade, M., Behlke, J., Lowenhaupt, K., Herbert, A., Rich, A., Oschkinat, H. A 6 bp Z-DNA hairpin binds two Z alpha domains from the human RNA editing enzyme ADAR1. *FEBS Lett.* **458**, (1) 27-31 (1999b).

Schroth, G.P., Kagawa, T.F. & Ho, P.S. Structure and thermodynamics of nonalternating C.G base pairs in Z-DNA: the 1.3-A crystal structure of the asymmetric hexanucleotide d(m5CGGGm5CG).d(m5CGCCm5CG). *Biochemistry* **32**, 13381-13392 (1993).

Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A. & Rich, A. Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841-1845 (1999a).

Schwartz, T., Lowenhaupt, K., Kim, Y.G., Li, L., Brown B.A. 2nd, Herbert, A., Rich, A. Proteolytic dissection of Zab, the Z-DNA-binding domain of human ADAR1. *J. Biol. Chem.* **274**, (5) 2899-2906 (1999b).

Sekimata, M., Takahashi, A., Murakami-Sekimata, A. & Homma, Y. Involvement of a novel zinc finger protein, MIZF, in transcriptional repression by interacting with a methyl-CpG binding protein, MBD2. *J. Biol. Chem.* (2001).

Shewchuk, B.M. & Hardison, R.C. CpG islands from the alpha-globin gene cluster increase gene expression in an integration-dependent manner. *Mol. Cell Biol.* **17**, 5856-5866 (1997).

Shlyapnikov, S.V., Lunin, V.V., Perbandt, M., Polyakov, K.M., Lunin, V.Y., Levdikov, V.M., Betzel, C., Mikhailov, A.M. Atomic structure of the Serratia marcescens

endonuclease at 1.1 A resolution and the enzyme reaction mechanism. *Acta Crystallogr D Biol Crystallogr.* May;56 (Pt 5):567-72 (2000).

Simpson, R.T. Nucleosome positioning: occurrence, mechanisms, and functional consequences. *Prog Nucleic Acid Res Mol Biol.*;40:143-84. Review (1991).

Simpson, R.T. Structure of chromatin containing extensively acetylated H3 and H4. *Cell,* 13: 691-699 (1978).

Singal, R., Ferris, R., Little, J.A., Wang, S.Z., Ginder, G.D. Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells. *Proc Natl Acad Sci U S A.* Dec 9;94(25):13724-9 (1997).

Singal, R., Wang, S.Z., Sargent, T., Zhu, S.Z. & Ginder, G.D. Methylation of promoter proximal transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. *J. Biol. Chem.* (2002).

Smith S.S., Kan, J.L., Baker, D.J., Kaplan, B.E., Dembek, P. Recognition of unusual DNA structures by human DNA (cytosine-5) methyltransferase. *J Mol Biol.* Jan 5;217(1):39-51 (1991).

Stancheva, I. & Meehan, R. Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev.* 14, 313-327. (2000).

Stancheva, I., Hensey, C., Meehan, R.R. Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *EMBO J.* 20(8): 1963-73. (2001).

Strauss, E.C., Andrews, N.C., Higgs, D.R. & Orkin, S.H. In vivo footprinting of the human alpha-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol. Cell Biol.* **12**, 2135-2142 (1992).

Takeuchi, H., Hanamura, N. & Harada, I. Structural specificity of peptides in Z-DNA formation and energetics of the peptide-induced B-Z transition of poly(dG-m5dC). *J. Mol. Biol.* **236**, 610-617 (1994).

Takusagawa, F. The crystal structure of d(GTACGTAC) at 2.25 A resolution: are the A-DNA's always unwound approximately 10 degrees at the C-G steps? *J Biomol Struct Dyn.* Feb;7(4):795-809 (1990).

Tate, P., Skarnes, W., Bird, A. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat Genet.* Feb;12(2):205-8. (1996).

Thoma, F. Nucleosome positioning. *Biochim Biophys Acta.* Feb 28;1130(1):1-19. Review. (1992).

Thomas, T.J. & Messner, R.P. Structural specificity of polyamines in left-handed Z-DNA formation. Immunological and spectroscopic studies. *J. Mol. Biol.* **201**, 463-467 (1988a).

Thomas, T.J., Baarsch, M.J. & Messner, R.P. Immunological detection of B-DNA to Z-DNA transition of polynucleotides by immobilization of the DNA conformation on a solid support. *Anal. Biochem.* **168**, 358-366 (1988b).

Thomas, T.J., Gunnia, U.B. & Thomas, T. Polyamine-induced B-DNA to Z-DNA conformational transition of a plasmid DNA with (dG-dC)n insert. *J. Biol. Chem.* **266**, 6137-6141 (1991).

Thorvaldsen, J.L., Duran, K.L., Bartolomei, M.S. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of the H19 and Igf2. *Genes Dev.* Dec 1;12(23):3693-702 (1998).

Tippin D.B., Sundaralingam M. Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCmCGG) and d(CCGGGCC(Br)5CGG) in three different conformations: Effects of Spermine binding and methylation on the bending and condensation of A-DNA. *J.Mol. Biol.* 267, 1171-1185 (1997a).

Tippin D.B., Ramakrishnan B., Sundaralingam M. Methylation of the Z-DNA decamer d(CG)5 potentiates the formation of A-DNA: Crystal Structure of d(Gm5CGm5CGCGCGC). *J. Mol. Biol.*, 270, 247-258 (1997b).

Tschiersch, B., Hofmann, A., Krauss, V., Dom, R., Korge, G., Reuter, G. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* Aug 15;13(16):3822-31 (1994).

Urnov, F.D. & Wolffe, A.P. Above and within the genome: epigenetics past and present. *J. Mamm. Gland Neoplasia.* Review. Apr, 6(2): 153-167 (2001).

Van Holde, K., Zlatanova, J. Unusual DNA structures, chromatin and transcription. Review. *BioEssays*, 16 (1), 59-66 (1994).

Vandel, L., Nicolas, E., Vaute, O., Ferreira, R., Ait-Si-Ali, S., Trouche, D. Transcriptional repression by the retinoblastoma protein through the recruitment of a histone methyltransferase. *Mol Cell Biol.* Oct;21(19):6484-94 (2001).

Van den Veyver, I.B., Zoghbi, H.Y. Methyl-CpG-binding protein 2 mutations in Rett syndrome. *Curr. Opin. Genet. Dev.* Jun; 10(3):275-9. Review. (2000).

Vargason, J.M., Eichman, B.F. & Ho, P.S. The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nat. Struct. Biol.* **7**, 758-761 (2000).

Vargason, J.M., Henderson, K., Ho, P.S. A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U S A* Jun 19;98(13):7265-70. (2001).

Verdaguer, N., Aymami, J., Fernandez-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J., Subirana, J.A. Molecular structure of a complete turn of A-DNA. *J. Mol. Biol.* 221, 623-635 (1991).

Venkataraman, S. Histone Acetylation and Nucleosome Dynamics. *PhD thesis*. The University of Edinburgh. (2001).

Voet, D. & Voet, J. G. *Biochemistry* Wiley, New York (1990).

Vorlickova, M., Sagi, J. Transitions of poly (dI-dC), poly (dI-methyl5dC) and poly (dI-bromo5dC) amoung and within the B-, Z-, A- and X-DNA families of conformations. *N.A.R.* 19, (9), 2343-2347 (1991).

Vyas, P., Vickers, M.A., Simmons, D.L., Ayyub, H., Craddock, C.F., Higgs, D.R. Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell* **69**, (5) 781-793 (1992).

Waddington, C.H. The strategy of the genes. London: Allen & Unwin (1957).

Wade, P.A., Jones, P.L., Vermaak, D., Wolffe, A.P. A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol.* Jul 2;8(14):843-6 (1998).

Wade, P.A., Wolffe, A.P. ReCoGnizing methylated DNA. *Nat Struct Biol.* Jul;8(7):575-7 Review. (2001).

Wade, P.A. Methyl CpG-binding proteins and transcriptional repression. *Bioessays* vol. 23, no. 12 Dec;23(12):1131-7. Review (2001).

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubrey, F., Wolffe, A.P. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation [see comments]. *Nat. Genet.* **23**, (1) 62-66 (1999).

Wakefield, R.I., Smith, B.O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A.P., Barlow, P.N. The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.* **291**, (5):1055-1065 (1999).

Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature.* Dec 13;282(5740):680-6 (1979).

Wang, Y., Thomas, G.A., Peticolas, W.L. Sequence dependent conformations of oligomeric DNA's in aqueous solutions and in crystals. *J Biomol Struct Dyn.* Oct;5(2):249-74 (1987).

Waterhouse, P.M., Wang, M.B., Finnegan, E.J. Role of short RNAs in gene silencing. *Trends Plant Sci.* Jul;6(7):297-301 (2001).

Watson, J.D.&.Crick, F.H.C. Molecular structure of nucleic acids. *Nature* **171**, 737-738. (1953); *and* Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171,** 964-967 (1953).

Weatherall, D.J. Phenotype-Genotype relationships in monogenic disease: Lessons from the thalassaemias. *Nature Reviews* **2**, 245-255. (2001).

Weston, S.A., Lahm, A., Suck, D. X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 A resolution. *J Mol Biol.* Aug 20;226(4):1237-56 (1992).

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R.E. Crystal structure analysis of a complete turn of B-DNA. *Nature* 287, 755-758 (1980).

Wittig, B., Dorbic, T. & Rich, A. Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei [published erratum appears in *Proc Natl Acad Sci U S A* 1991 Aug 1;88(15):6898]. *Proc. Natl. Acad. Sci. U. S. A* **88**, 2259-2263 (1991).

Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W. & Rich, A. Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.* Dec;11(12): 4653-63 (1992).

Wolffe, A.P. Nucleosome positioning and modification: chromatin structures that potentiate transcription.*Trends Biochem Sci.* Jun;19(6):240-4. Review (1994).

Wu, H.M. & Crothers, D.M. The locus of a sequence-directed and protein-induced DNA bending. *Nature.* 308(5959): 509-513 (1984).

Wu, C. & Morris, J.R. Genes, Genetics, and Epigenetics: A Correspondence. *Science* **293**, 1103-1105 (2001).

Xiang, F., Buervenich, S., Nicolao, P., Bailey, M.E., Zhang, Z., Anvret, M. Mutation screening in Rett syndrome patients. *J. Med. Genet.* Apr;37(4):250-5 (2000).

Xin, Z., Allis, C.D., Wagstaff, J. Parent-specific complementary patterns of histone H3, lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. *Am. J. Hum. Genet.* Dec;69(6):1389-94 (2001).

Yagi, M., Gelinas, R., Elder, J.T., Peretz, M., Papayannopoulou, T., Stamatoyannopoulos, G., Groudine, M. Chromatin structure and developmental expression of the human alpha-globin cluster. *Mol Cell Biol.* Apr;6(4):1108-16. (1986).

Yamaki, H., Ohtsubo, E., Nagai, K., Maeda, Y. The oriC unwinding by dam methylation in Escherichia coli. *Nucleic acids Res.* Jun 10;16(11):5067-73 (1988).

Yang, J., Camakiaris, H., Pittard, A.J. Further genetic analysis of the activation function of the TyrR regulatory protein of Escherichia coli. *J. Bacteriol.* 178(4): 1120-5 (1996).

Yenidunya, A., Davey, C., Clark, D., Felsenfeld, G., Allan, J. Nucleosome positioning on chicken and human globin gene promoters in vitro. Novel mapping techniques. *J. Mol. Biol.* Apr 8;237(4):401-14 (1994).

Yoon, C., Prive, G.G., Goodsell, D.S., Dickerson, R.E. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc Natl Acad Sci U S A.* Sep;85(17):6332-6 (1988).

Yu, F. & Schumann,G.S. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* 29(21), 4493-4501. (2001).

Yusufzai, T.M., Wolffe, A.P. Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res.* Nov 1;28(21):4172-9 (2000).

Zacharias, W., O' Connor, T.R., Larson, J.E. Methylation of cytosine in the 5-position alters the structural and energetic properties of the supercoil-induced Z-helix and of B-Z junctions. *Biochemistry*, 27; 2970-2978 (1988).

Zacharias, W., Jaworski, A. & Wells, R.D. Cytosine methylation enhances Z-DNA formation in vivo. *J. Bacteriol.* **172**, 3278-3283 (1990).

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., Reinberg, D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13**, (15) 1924-1935 (1999).

Zhang, Y.& Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 15, 2343-2360 (2001).

Zhou, N., Bianucci, A.M., Pattabiraman, N., James, T.L. Solution structure of [d(GGTATACC)]2: wrinkled D structure of the TATA moiety. *Biochemistry.* Dec 1;26(24):7905-13 (1987).

Zhou, G.W., Ho, P.S. Stabilization of Z-DNA by demethylation of thymine bases: 1.3-A single-crystal structure of d(m5CGUAm5CG). *Biochemistry.* Aug 7;29(31):7229-36 (1990).

Zhou, N., Manogaran, S., Zon, G., James, T.L. Deoxyribose ring conformation of [d(GGTATACC)]2: an analysis of vicinal proton-proton coupling constants from two-dimensional proton nuclear magnetic resonance. *Biochemistry.* Aug 9;27(16):6013-20 (1988).

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M., Jost, J.P. 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucl. Acid. Res.* 28(21)4157-65. (2000).